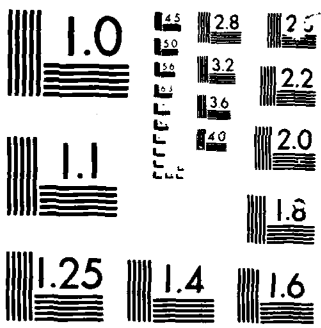
AD-A167 931 A STUDY TO DETERMINE THE RELATIVE SKILL OF FOUR MODEL OUTPUT STATISTICS P..(U) NAVAL POSTGRADUATE SCHOOL MONTEREY CA S J FATJO MAR 86 1/1

UNCLASSIFIED F/G 9/2 NL



MICROCOPY CHART

AD-A167 931

DTIC FILE COPY

NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

A STUDY TO DETERMINE THE RELATIVE SKILL OF FOUR MODEL OUTPUT STATISTICS PREDICTION METHODS USING SIMULATED DATA FIELDS

by

Steve J. Fatjo

March 1986

Thesis Advisor: R. J. Renard
Thesis Co-advisor: R. W. Preisendorfer

Approved for public release; distribution is unlimited

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a REPORT SECURITY CLASSIFICATION	1b. RESTRICTIVE MARKINGS
2a SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.
2b DECLASSIFICATION/DOWNGRADING SCHEDULE	
4 PERFORMING ORGANIZATION REPORT NUMBER(S)	5 MONITORING ORGANIZATION REPORT NUMBER(S)
6a NAME OF PERFORMING ORGANIZATION: Naval Postgraduate School 6b OFFICE SYMBOL (If applicable): Code 63	7a NAME OF MONITORING ORGANIZATION: Naval Postgraduate School
6c ADDRESS (City, State, and ZIP Code): Monterey, California 93943-5100	7b ADDRESS (City, State, and ZIP Code): Monterey, California 93943-5100
8a NAME OF FUNDING/SPONSORING ORGANIZATION 8b OFFICE SYMBOL (If applicable)	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c ADDRESS (City, State, and ZIP Code)	10 SOURCE OF FUNDING NUMBERS: PROGRAM ELEMENT NO / PROJECT NO / TASK NO / WORK UNIT ACCESSION NO

11 TITLE (Include Security Classification)

A STUDY TO DETERMINE THE RELATIVE SKILL OF FOUR MODEL OUTPUT STATISTICS PREDICTION METHODS USING SIMULATED DATA FIELDS

12 PERSONAL AUTHOR(S)

Fatjo, Steve J.

13a TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
Master's Thesis	FROM ____ TO ____	1986 March	72

16 SUPPLEMENTARY NOTATION

17 COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	ANOVA; Forecast; Model Output Statistics Prediction Methods; Simulated Data Fields

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

This report describes the testing of four Model Output Statistics prediction mehtods on simulated data fields for the purpose of determining their relative skills in forecasting a generic weather parameter (predicthand). Of the four methods, three use Bayes Law of Inverse Probability to discriminate, while the other method uses conditional Proba. ility. The simulated data sets, models and observers necessary to accomplish this goal are created according to a uniquely developed simulation design. The results indicate that there is a definite diffusing conditional probability, to forecast the weather parameter. Through the use of Analysis of Variance (ANOVA) technique, this difference is found to be significant with respect to chance.

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a NAME OF RESPONSIBLE INDIVIDUAL R. J. Renard	22b TELEPHONE (Include Area Code): 408-646-2516 22c OFFICE SYMBOL: 63Rd

DD FORM 1473, 84 MAR

83 APR edition may be used until exhausted

All other editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

Approved for public release; distribution is unlimited.

A Study to Determine the Relative Skill of Four Model Output Statistics Prediction Methods Using Simulated Data Fields

by

Steve J. Fatjo
Lieutenant, United States Navy
B.S., Old Dominion University, 1979

Submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN METEOROLOGY AND OCEANOGRAPHY

from the

NAVAL POSTGRADUATE SCHOOL
March 1986

Author: Steve J. Fatjo
Steve J. Fatjo

Approved by: Robert J. Renard
Robert J. Renard, Thesis Advisor

Rudolph W. Preisendorfer
Rudolph W. Preisendorfer, Co-Advisor

Robert J. Renard
Robert J. Renard, Chairman,
Department of Meteorology

J. Dyer
John N. Dyer,
Dean of Science and Engineering

ABSTRACT

This report describes the testing of four Model Output Statistics prediction methods on simulated data fields for the purpose of determining their relative skills in forecasting a generic weather parameter (predictand). Of the four methods, three use Bayes Law of Inverse Probability to discriminate, while the other method uses conditional probability. The simulated data sets, models and observers necessary to accomplish this goal are created according to a uniquely developed simulation design. The results indicate that there is a definite difference in the ability of one of the four methods, namely the method using conditional probability, to forecast the weather parameter. Through the use of the Analysis of Variance (ANOVA) technique, this difference is found to be significant with respect to chance.

A-1

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I wish to thank the following members of the Model Output Statistics project team for their assistance in the preparation of this thesis: Robert Hale, computer programming consultant and "corporate knowledge" of the MOS prediction methods, for his invaluable help in keeping the methods straight and in programming the simulation process; Mr. Robert Godfrey, Naval Environmental Prediction Research Facility (NEPRF), MOS project leader, for his helpful comments; Mr. Paul Lowe, NEPRF, statistician, method developer and consultant, for his insightful ideas on how to analyze the results using ANOVA; Dr. Rudolph Preisendorfer, thesis co-advisor, method developer, statistical theoretician, and consultant, for his brilliant ideas on how to do a simulation study, and for his thoughful guidance in all phases of the thesis effort; and Prof. Robert Renard, thesis advisor and NPS MOS project head, for his patience and steadfast support all the way. In addition to the members of the MOS project team, I wish to thank my dear wife, Sherry, and my children, for their love and understanding during the concluding months of this thesis effort. Above all, I wish to thank the Lord, my God, without whom none of this would have been possible.

I. INTRODUCTION

The Model Output Statistics (MOS) approach to forecasting consists of relating the numerical model output parameters, diagnostic and prognostic (predictors), to sensible operationally-important weather parameters (predictands), e.g. visibility, cloud amount, precipitation, for the purpose of enhancing the skill of forecasting these parameters (Glahn and Lowry, 1972).

The first major MOS work was readied for application in the early 1970's by the National Weather Service (NWS), the weather arm of the National Oceanic and Atmospheric Administration (NOAA). More recently, both the U.S. Navy and the U.S. Air Force have been involved in the development of MOS schemes. NOAA's programs are operational, both at civilian and military sites, with continuing development by the Techniques Development Laboratory (TDL), National Weather Service, Silver Springs, MD.

The Navy's MOS effort began at the Naval Postgraduate School (NPS) in Monterey, CA in the mid/late 1970's, using visibility and fog over the North Pacific Ocean as the predictands of interest (Renard and Thompson, 1984; Koziara, Renard and Thompson,1983). These experiments were limited and the results not immediately operationally applicable. However, the studies formed the basis for a decision by the U.S. Navy to pursue the development of MOS for all oceans of the world, for a select number of air/ocean parameters. Following this decision in 1981, a series of studies were initiated at NPS, as a joint effort of NPS's Meteorology Department and the Naval Environmental Prediction Research Facility (NEPRF), Monterey, CA The first study (Karl,1984) investigated the use of three conditional probability MOS

prediction schemes developed by Preisendorfer[1] (1983a,b), using visibility at the model initialization time as the predictand and the output from Fleet Numerical Oceanography Center's (FNOC) Navy Operational Global Atmospheric Prediction System (NOGAPS) model as predictors, applied to a limited homogeneous region of the North Atlantic Ocean. Karl intercompared the results of these three methods and multiple linear regression methods with variable thresholding, as proposed by Lowe (1984). Diunizio (1984) followed Karl with a similar experiment, but for additional homogeneous Atlantic Ocean areas, and forecast intervals to 48 hours, and with modifications to the MOS methodologies. The third study (Wooster, 1984) concentrated on cloud amount and ceiling, using essentially the same MOS methods as his predecessors with further variations in the multiple linear regression threshold model. The most recent effort (Elias, 1985), contrasted a new model, namely the Principal Discriminant Method (Preisendorfer, 1984), with the earlier methods on their ability to predict visibility.

This study concerns the testing of three MOS prediction methods exercised by the previous NPS investigations (the Maximum-Probability Method II, the Multiple Linear Regression Method, and the Principal Discriminant Method), plus one additional method (Discriminant Analysis Method), on statistically-derived simulated (i.e., controlled) predictor/predictand data sets, with the goal of ranking the methodologies as to their relative skill in predictand specification.

---

[1]Dr. R.W. Preisendorfer was the Naval Air Systems Command G. J. Haltiner Research Chair Professor in the Department of Meteorology, Naval Postgraduate School, Monterey, CA for 1983. Dr. Preisendorfer is currently affiliated with NOAA's Pacific Marine Environmental Laboratory (PMEL) in Seattle, WA.

II. OBJECTIVES AND APPROACH

There are three main objectives in this particular study contributing to the U.S. Navy's MOS program. The first objective is to develop fields of simulated data that have controllable predictor/predictand parameters, such that the conditions of predictability of the data fields can be varied and the results made reproducible. The second objective is to test four MOS prediction methods, using the simulated data fields, in order to determine their relative ranking in the skill of predicting the generic sensible weather parameter. The third objective is to determine the most skillful MOS method on which further testing could be concentrated.

The approach taken to fulfill these objectives was to program the simulation procedures outlined by Preisendorfer (1985) so as to create the data fields. A training and three test sets were generated for each MOS methodology in order to provide ample scoring statistics upon which to analyze the results. For this study, two such training/testing sets of simulated data were generated with 1200 rows of nine columns, each column representing a real primary data field. One set was called 'the easy data set', while the other was called 'the hard data set'. The easy data set was one that the MOS prediction methods could easily make a prediction from, due to the relatively high correlation between the predictors and the predictand. The hard data set was one that would have less correlation between the predictors and the predictand, thus making a prediction more difficult. The two data sets were determined in such a way as to have a signal-to-noise ratio[2] of 4:1 for the easy data set and 1:1

[2]The signal-to-noise ratio is defined as the maximum ratio of the between-class distance to the within-class

for the hard data set. Two numerical weather prediction models were also simulated in this study. The first is called the good model because it introduced less distortion and less noise to the signal, while the bad model produced a greater distortion and a larger noise component. Also in this study, three observers were simulated in order to look at the resultant variation in predictive skill of the four methods, as the skill of the observer varied. The three observers were: i) the perfect observer - one who never makes a wrong observation of the actual weather event; ii) the good observer - one who occasionally makes a wrong observation of the actual weather event, but not by more than one category; and lastly iii) the bad observer - one who makes a wrong observation of the actual weather event more often than the good observer, and sometimes by more than one category.

---

distance of two (or more) multivariate populations given in the function:

$$S/N = [\mu(1)-\mu(2)]2/\sigma 2 \quad \text{(for univariate situation),}$$

where $\mu(1)$ is the mean of the first category of the three being compared, $\mu(2)$ is the mean of the other category and σ^2 is the average variance of the two categories being compared.

III. METHODOLOGIES OF THE MOS SIMULATED DATA SETS

The sections that follow describe the procedures used to create the simulated data sets, the simulated models and the simulated observers for this study, as based on guidance provided by Preisendorfer (1985).

A. THE MOS SIMULATION PROBLEM

The problem of designing an MOS simulation process is threefold in nature. Besides designing the simulated natural data fields, there is also the problem of prescribing the skill of the observer viewing them, and of defining how accurately a numerical weather prediction model would reproduce them. The observer, on the one hand, would be viewing a parameter, such as visibility at sea, and have to estimate it subjectively. On the other hand, the model will be reproducing output parameters such as pressure, temperature, winds, etc., that could be used by any of the MOS prediction methods to predict the visibility, or other sensible weather parameters.

In setting up an MOS scheme, i.e., to train it to forecast the sensible weather parameter, the observations of the predictand must be coupled with the output parameters of the model. Then, when the method is to be tested, it will, according to its training, take a fresh model predictor set and produce a forecast of the predictand. Thus, in attempting to design a procedure using the simulated data sets, where there is built - in correlation between the real predictors, a simulation of the model fields must be done as well, including whatever biases and errors that are inherent to numerical weather prediction models. The ability of the observer must also be modeled so that it does a controllably imperfect job in estimating the weather parameter in question. This idea is illustrated in Fig. 1.

The left-hand column represents the real atmosphere or ocean as given to us by nature. The real primary fields are those which are normally measured or are known in principle. Specifically, the real primary fields are those which are normally measured or observed and incorporated routinely into the Newtonian equations of motion and the Laws of Thermodynamics. The real secondary fields are those of forecasting interest in this study, since the numerical models do not usually forecast them directly. The right-hand column represents the modelled atmosphere or ocean. The model primary and secondary fields here are only as good as the model that derives them and the human that observes them, respectively. It is recognized that there is always some level of error in these fields. The MOS methods in this simulation study will take the output parameters (predictors) from the modelled primary field, pair them with the simulated observations of the secondary field, and use these pairings subsequently to forecast the simulated real secondary field parameter (predictand).

B. THE MOS PREDICTION CONCEPT

Now for a more detailed look at the MOS procedure outlined above. A generalization of the MOS approach to prediction of sensible weather parameters is shown in Fig. 1. Some estimated value of a real secondary field parameter (such as visibility) is recorded by an observer for some location at time t. This estimate is called the estimated predictand since the secondary field is the field that is to be predicted using statistics derived from these estimates. Meanwhile, at the same location and time t the numerical weather prediction model produces values for the model primary fields. The model primary fields are called the model predictors.

These model predictor/estimated predictand pairings are taken over a large region of interest and provide what shall

be termed the Basic Data Set (BDS). An MOS method (X, Y, ..., Z) is then trained on a part of the BDS to forecast a value of the predictand field from the set of model predictors it is given. This method is then tested on the remaining part of the BDS as follows.

1) A set of model predictors is chosen from the testing part of the BDS, using the predictors obtained during the training stage.
2) The method then produces a forecast of the predictand from this set.
3) This forecast predictand is then compared with the estimated predictand that was paired with the set of model predictor values used.
4) Then a skill score is assigned to the forecast of this particular method.
5) This procedure is then repeated for the other MOS methods to be tested.

Once all of the scores have been attained, the various methods are intercompared using these skill scores and some assessment is made as to their relative abilities to accurately predict the estimated predictand value from the given set of model predictors. The important assumption made here is that the methods will have the same relative ranking using the simulated data as they would have when subsequently tested with a fresh set of model predictor values from an actual data set of model output. This assumption will be well-founded provided that:

1) The simulated data set is representative (in a statistical sense) of the real primary and secondary fields encountered in nature; and if
2) The errors of the models and observers have been well-simulated.

C. SIMULATING THE REAL PRIMARY FIELDS

The real primary fields, which may be thought of as time series at p fixed points in space, were generated from a set of pxp covariance matrices using a formulation scheme suggested by Preisendorfer (1985, Appendix C). The field-to-field (i.e., cross-) correlations are represented by the

matrix M, while the time lagged field-to-field correlations are represented by pxp matrix K. The scheme was developed with the goal of saving on computation time by defining the covariance matrices to be associated with stationary processes in space. For example, suppose m(x,x') is the entry of M in row x and column x', then the matrices are stationary, in that the covariances depend only on the differences (x-x') of the arguments x and x' in m(x,x'), so that it may be written as m(x-x') and k(x-x') for K. In this context, the eigenvalues and eigenvectors of M and K are expressible using simple algebraic formulas. The need for various matrix manipulating subroutines is thereby eliminated.

The formula used to create the pxp M matrix is:

$$m(x,x') = \exp(-\mu \cdot |x-x'|) \tag{3.1}$$

$x,x' = 0, \ldots, p-1$ for $0 < \mu < \infty$, where μ is the variable that controls the field-to-field correlation. The pxp K matrix is formed similarly:

$$k(x,x') = \kappa_a \cdot \exp(-\mu \cdot |x-x'|) \tag{3.2}$$

$x,x' = 0, \ldots, p-1$ for $0 < \kappa_a < 1$, where κ_a is the variable that controls how much of a lag-induced difference there is between matrices M and K. (This is a simplified version of a more general approach in Preisendorfer (1985, Appendix B).) In this study, p was set to equal 9, so that the time series used simulated nine points in space.

For this study two separate data sets were formed. The difference between them is the amount of correlation and the amount of lag defined for the 9x9 matrices M and K. The degree to which the set of predictors produced by the model were correlated and also the amount of lag-effect between

the pair of matrices determined whether or not the MOS prediction method would have an easy or a hard time forecasting the sensible weather paramater (predictand). Table I gives an example of the covariance matrices M and K with μ = 0.15 and κ_a = 0.90. These values provide a set of covariance matrices that are well-correlated and thus labelled the 'easy' data set. The values for the variables of the 'hard' data set are μ = 0.50 and κ_a = 0.60. This produced a correlation that is much lower and a lag-induced effect that is greater. Therefore, the forecast made from this Basic Data Set is more difficult. The lower half of Table I shows these matrices from the hard data set.

The range of values of the x,x' pairs is especially tailored for the spatially stationary context, and in fact, arithmetic modulo p must be used on the spatial-index x-x' values that do not fall within the prescribed range. For example, if x-x' is not in the set {x: 0, ..., p-1}, x-x' must be reduced modulo p to map it into the finite set. The diagram in Fig. 2 illustrates this idea for the case p = 9. The range of x,x' can be visualized as being on a circle where the 0,..., p-1 (=8) values are plotted on it. Any values of x-x' outside their range are wrapped around the circle modulo p, so, in this way, all members of the set of integers (each representing the spatial location of a time series) can be handled.

By construction, the values of m(x,x') and k(x,x') depend only on x-x', and by symmetry we have m(x,x') = m(x',x) and k(x,x') = k(x',x). Hence, from now on m(x,x') = m(0,x-x') can be written as m(x-x'), i.e., with only one argument. In relating the diagram in Fig. 2 to the matrix values in Table I, notice that m(8,7) = m(1), while m(7,8) = m(-1) = m(8), and since, by symmetry, m(7,8) = m(8,7), it can be seen that m(8) = m(1). The same holds true for matrix K. Table II shows M and K in the simplified notation.

The data sets were generated using the following autoregressive equation:

$$Z(t,x) = 0.5 \cdot \nu(0) \cdot b(0,t) + \{\sum_{j=1}^{m} \nu(j) \cdot \tag{3.3}$$
$$[b(j,t) \cdot \cos(\kappa(j) \cdot x) + c(j,t) \cdot \sin(\kappa(j) \cdot x)],$$

where m=4 and p=2•m+1 for j=0, ..., m, with $t \epsilon J$ (the set of integers); x=0, ..., p-1; $\nu(j)=(\lambda(j)/p)^{1/2}$; $\kappa(j) = (2 \cdot \pi \cdot j)/p$.

The terms used in Eq.(3.3) will be elaborated on individually or in pairs in the following paragraphs, starting with the first equations derived from Eqs.(3.1) and (3.2).

The $\lambda(j)$ term found in the expression for $\nu(j)$ is the set of eigenvalues determined for the M matrix by the following formula:

$$\lambda(j) = 1.0 + 2.0 \cdot \{\sum_{x=1}^{m} m(x) \cdot \cos[\kappa(j) \cdot x]\} \tag{3.4}$$

with $\lambda(p-j) = \lambda(j)$ using modular arithmetic, and $\kappa(j) = (2 \cdot \pi \cdot j)/p$, for j = 1, ..., m. Through this variable the field-to-field matrix correlations were expressed.

The autogressive correlations $\rho(j)$ were given by the formula:

$$\rho(j) = k(0) + 2.0 \cdot \{\sum_{x=1}^{m} k(x) \cdot \cos[\kappa(j) \cdot x]\}/ \lambda(j) \tag{3.5}$$

for j = 0, ..., m; with $\rho(p-j) = \rho(j)$ for j = 1, ..., m again by modular arithmetic. This relation takes into account the lag correlation and is used to obtain the variances for the random forcing of Eq. (3.3). These variances were obtained through use of the formulae:

$$\sigma^2(\beta,j) = 2.0 \cdot (1 - \rho^2(j)) \cdot (1 + \delta(0,j)) \tag{3.6}$$

$$\sigma^2(\gamma,j) = 2.0 \cdot (1 - \rho^2(j)) \cdot (1 - \delta(0,j)) \qquad (3.7)$$

for $j = 0, \ldots, m$, where $\delta(0,j)$ is a special case of Kronecker's Delta function $\delta(i,j)$ with $i = 0$. In general, $\delta(i,j) = 1$ if $i = j$, and $= 0$ if $i \neq j$.

The square root of the variances obtained in Eqs.(3.6) and (3.7) (i.e., the standard deviations) are used in a random number generator subroutine[3] to get the random forcing terms, such that their values are normally distributed with a zero mean, i.e.:

$$\beta(j,t) \sim N(0,\sigma^2(\beta\ ,j)) \text{ and } \gamma(j,t) \sim N(0,\sigma^2(\gamma\ ,j)) \qquad (3.8)$$

Observe that $\sigma^2(\beta\ ,j) = \sigma^2(\gamma\ ,j)$ for $j = 1, \ldots, m$, and that $\sigma^2(\beta\ ,0) = 4 \cdot [1-\rho^2(0)]$, while $\sigma^2(\gamma\ ,0) = 0$.

The autoregressive correlations $\rho(j)$ are used again along with the random forcing terms to develop the time-dependent coefficients $b(j,t)$ and $c(j,t)$ via the following formulae:

$$b(j,t) = \rho(j) \cdot b(j,t-1) + \beta(j,t) \qquad (3.9)$$

$$c(j,t) = \rho(j) \cdot c(j,t-1) + \gamma(j,t) \qquad (3.10)$$

for $j = 0, \ldots, m$.[4]

D. SIMULATING THE REAL SECONDARY FIELD

The next step in the simulation process is the modelling of the real secondary fields (predictands). This required defining a link between the real primary and secondary

---

[3]The random number generator used is NPS's W.R. Church Computer Center's library subroutine, GGUBS, which creates variates with uniform distributions. These uniform distributions are changed into gaussian normal distributions through the use of the method of Box and Muller (1958).

[4]Notice that $c(0,t)$ is uniquely zero due to the fact that $\sigma^2(\gamma\ ,0) = 0$.

fields. The fact of the matter is that there are, in many cases, no readily available algorithms which can be applied to the real primary fields to obtain the real secondary fields. Indeed, this is at present a vigorously pursued set of problems in meteorology and oceanography. This fact is the basis for the need of MOS-type procedures in the first place! Fortunately, in this simulation study the real forms of such linkages are not needed. Therefore, in order to provide the linkage desired, it is sufficient to invent some reasonable-appearing relations. Therefore, predictors could have been combined together in any number of different ways, since the number of algebraic and analytic possibilities are endless; for the present study the method chosen is the following. There are in the set of real primary fields generated by Eq.(3.3), nine columns of predictor time series, Z(t,0) through Z(t,8), for t = 1, ..., 1200, where 1200 is the number of entries in each column. By taking one of these columns and relabelling it the real secondary field, the desired effect of having a linkage between the predictors and the predictands is assured.

So now, by following this procedure, there is one predictand column and eight predictor columns with a correlation factor and a lag factor that can be controlled by the use of the μ and the κ_a terms. This method of generating predictors and predictand is termed the <u>In-House Field Method</u>, see (Preisendorfer,1985; Appendix B).

For this study, the predictand column chosen is the Z(t,0) - column because of the symmetry of the values to either side of it. The correlation values are highest for the nearest two neighbors, Z(t,1) and Z(t,8), and decrease in order going away from Z(t,0) in either direction. The correlation of the eight predictor columns to the predictand column is shown in Fig. 3 and Fig. 4 for the easy and hard data sets, respectively.

Next, the predictand column is sorted by entry magnitudes, with the entries of lowest magnitude being placed at the top of the column, while those rows with the largest magnitudes are placed at the bottom. At the same time, the other eight columns of predictors are also sorted. In this way the row relationship between the predictand (Z(t,0) value) and the predictors (Z(t,1) through Z(t,8)) is maintained. Once this is accomplished, the values in the predictand column are grouped in the following manner.

1) For the upper 400 entries, those with the lowest numerical values are tagged with the value 1 and comprise category 1 (analogous to forming the 'below' tercile in meteorological practice).

2) For the middle 1700 entries, those between the two extremes are tagged with the value 2 and comprise category 2 (forming thereby the 'normal' tercile).

3) For the lowest 400 entries, those with the largest numerical values are tagged with the value 3 and comprise category 3 (forming the 'above' tercile).

The middle category, category 2, contain the most entries because of a desire to keep the variances for the three categories equal, or nearly so. This proved to be quite a challenge, and only after many experiments with various category sizes was it accomplished. Fig. 5 shows an example of the three categories with equally populous intervals (400 entries in each). The high narrow spike in the middle category indicates that the variance for that category was much less than either of the two categories on the wings, and a small variation between the training set and the test sets led to widely different verification scores. This result was unsatisfactory, and so a different interval size was sought that would not lead to this type of instability. The interval size that was settled on[5] was to have the interval size determined by the standard deviation.

---

[5] This idea of interval spacing using the standard deviation as a measure was suggested by Lowe in a private discussion, and upon testing proved to have the stability desired when there were at least 400 entries per predictor column.

The overall interval spacing for a gaussian (normal) distribution can be defined by ± three standard deviations either side of the mean. Hence, for this study, the middle two standard deviation (sigma) intervals are defined as the middle category, category 2, while the two outer sigma intervals on either side are defined as category 1 and category 3 as shown in Fig. 6.

The final step for the predictand-category defining process is to randomly sample down from the 1700 entries in the middle category to obtain the desired 1200 entry size for each column, which means that each category subset has 400 values, and also nearly the same variance.

E. SIMULATING THE MODEL PRIMARY FIELDS (MODEL PREDICTORS)

The counterparts to the real primary fields are the model primary fields, as shown in Fig. 1. The model primary fields are the imperfect versions of nature's real primary fields since they contain some distortion and noise due to the inability of man to model the atmosphere and ocean accurately.

The model imperfection is simulated in this study through the following equation:

$$X(t,x) = \sum_{x'=1}^{p-1} S_T(x,x') \bullet Z(t,x') + n(t,x). \qquad (3.11)$$

Here the model primary field ($X(t,x)$) is produced first by having the real primary field ($Z(t,x)$) distorted through multiplication with a matrix $S_T(x,x')$ that consists of a fraction on the diagonal and zero elsewhere, and second by having some noise added to it, element by element. The S_T matrices used in this simulation are shown in Table III. They are special cases of the more general linear transformations possible on the $Z(t,x)$ field.

The value on the diagonal for the good model is 0.95 and for the bad model; it is 0.50. The noise term $n(t,x)$ is

created using the same random number generator as before, only now different values are used for the variances to control the amount of spread about the zero mean. The normal distributions obtained are generated with $\sigma^2 = 0.01$ for the good model, and with $\sigma^2 = 0.25$ for the bad model. Thus, the good model has 95% of the original Z(t,x) value, plus a random perturbation from the centered normal distribution with standard deviation $\sigma = 0.1$. The bad model has only 50% of the original Z(t,x) value, and a perturbation of standard deviation $\sigma = 0.5$.

F. SIMULATING THE OBSERVED REAL SECONDARY FIELDS (ESTIMATED PREDICTAND)

The final simulation requires the creation of an observer who will make estimates of the real secondary field (estimated predictand). These observations with their inherent errors will be combined with the model predictors by the MOS prediction methods and be used to forecast the real secondary field parameter as shown in Fig. 1.

For this study, the estimated predictand was created with marine atmospheric visibility in mind as the sensible weather parameter being forecast. So the estimates of the predictand are grouped into discrete categories according to whatever limits are desired to separate the categories. In this study the generic predictand (considered as marine atmospheric visibility) was grouped into three categories - 1, 2, 3, representing good, marginal and bad visibility, respectively. In making an estimation of the predictand, the observer may correctly choose the category of the actual event, or the observer may miss and choose one of the other two categories. For example, if there is only marginal visibility occurring at some place, for some time t, then the observer can choose category 2, and be correct, or category 1 or 3, and be wrong. A good observer would have the ability (or skill) to make the correct estimate most of the

time, but a bad observer wouldn't. Table IV shows a 3x3 table which summarizes the relative frequency e(i,j) of the observer estimating category i when, in fact, category j occurs in nature. The e(i,j) values are normalized so that each of the columns sum to 1. Here a perfect observer would have e(i,j) = 0 when i ; j, and e(i,j) = 1 when i = j, for i = 1,2,3. A bad observer would have greater off-diagonal values indicating an inability to observe correctly.

The observer's estimate is modeled in the following manner. The real secondary field (predictands), described earlier in Section D, are arranged in ascending order according to their numerical values, and relabelled 1, 2, 3, such that the upper 400 values equal 1, the middle 400 values equal 2, and the last 400 values equal 3. To simulate the observer's role a uniformly distributed number u is chosen on the interval I = [0,1] which has been partitioned according to the observer's skill, i.e.:

$$\begin{aligned} A(1,j) &= \{ u: 0 \le u \le e(1,j) \} \\ A(2,j) &= \{ u: e(1,j) < u \le e(1,j) + e(2,j) \} \\ A(3,j) &= \{ u: e(1,j) + e(2,j) < u \le 1 \}, \end{aligned} \tag{3.12}$$

so that I = A(1,j) + A(2,j) + A(3,j), j = 1, 2, 3. For example, if the number u, chosen randomly, falls into A(2,j), then the category assigned will be category 2 for the estimated predictand when in fact category i,j = 1,2,3 occurs.

The three observers simulated in this study have skills[6] decreasing from 100% for the perfect observer, to 87% for the good observer, and still lower to 69% for the bad observer. Their respective skills are shown in Table V.

---

[6] By skills it is meant the ability to observe correctly the actual weather event which is occurring at the time of observation. This ability can be defined by the calculation of the average of the main diagonal elements of the verifying relative frequency (i.e., contingency) table.

IV. THE MOS PREDICTION METHODS AND THEIR RESULTS

A. METHOD DESCRIPTIONS

1. Principal Discriminant Method

The Principal Discriminant Method (PDM) was proposed by Preisendorfer (1984). This method is basically a discriminant method in that the data are partitioned into predictand classes and forecasts are made by forming probabilities for the predictor as to whether it belongs to one predictand class or the other. This particular method of discrimination is distinguished by the fact that it fits a gaussian probability density distribution (for example) to each predictand category subset, using a principal component analysis of the data points in the category subset. The method's most novel feature is that, if the categorical distributions are significantly non-gaussian, then a successive, controlled splitting of the category subset is performed in the local principal component coordinate frame to obtain a better fitting of probability density functions. With each test set, fresh values of predictors are used and the associated probability density values for each category are found. The forecast of the predictand is then made using Bayes Law of Inverse Probability.

The Principal Discriminant Method also contains a methodology for predictor selection. Although each of the four methods to be tested has its own method of predictor selection, the PDM was chosen to select the three predictors to be used by all of the methods. The basic groundwork for the PDM and its variable-selection process was programmed by Elias (1985) in three separate programs (a single predictor screening program, a predictor correlation program and a

multiple predictor program). Hale[7] subsequently made additional modifications to these programs to more closely follow Preisendorfer's original formulation.

The procedure used to select variables and implement the PDM is now as follows.

1) The initial screening program is run where each predictor is separately ranked according to potential predictability (PP) and potential total percentage correct (PAO) for the training set. The first predictor chosen is that with the highest PP.

2) Then the correlation program is run to find the next candidate predictor to be added such that it is the least correlated to those already chosen.

3) The PDM multiple predictor program which implements the methodology described above is then run on these predictors chosen in step 2. If the PP and PAO scores both increase then the candidate predictor chosen in step 2 is added. If the PP and PAO do not increase then step two is repeated. The process is carried out until three predictors are chosen.

4) With the three chosen predictors, the PDM multiple predictor program is run using the test sets. The program compiles a contingency table and computes verifying statistics.

2. Maximum-Probability Method II

The Maximum-Probability Method II (MaxProb2) was also proposed by Preisendorfer (1983a,b) and exercised in all of the previous MOS studies. It differs from the discriminant methods in that the forecasts are based on conditional probabilities for given values of the predictor. To accomplish this, the data are first classed into a n-dimensional predictor space. The discrete cells into which the points are placed are of a size determined by dividing each predictor into equally populous intervals. Then within each cell the number of points belonging to each predictand class is tallied and the conditional probabilities for each predictand class are formed. Thus, each specific cell has its own conditional probabilities and the predictand

[7] Robert A. Hale, NPS, joined the NPS/NEPRF MOS project team in February 1985. He has principal responsibility for the management of the MOS-archived data fields and Fortran programs.

category which has the maximum conditional probability is forecast for that cell. With each test set, fresh values of predictors are used to identify the cell location in the predictor space and the forecasted predictand category is returned. A contingency table of forecast versus observed values is then formed from which the verifying statistics can be calculated.

The procedure used in this study employed two Fortran programs written by Karl (1984).

1) First, a program is run which determines the number of equally populous intervals into which each predictor should be divided.

2) A program which implements the Maximum-Probability methodology for a multiple number of predictors is run second. In this study, the three predictors were chosen by the Principal Discriminant Method.

3. Discrimination by Dimension Reduction using Regression

The Discrimination by Dimension Reduction using Regression (DDRR) method was proposed by Lowe.[8] This method, in combination with various thresholding techniques, has been used in all previous NPS MOS studies since Karl (1984), where it was called the Multiple Linear Regression method. The method uses the BMDP Statistical Software [University of California, 1983] programs - P1R, P5D and P4F. The P1R program carries out a multiple linear regression on the input predictor distributions; The P5D program displays various statistics (mean and variance of the classification functions) and histograms for the dimensionally-reduced estimated variable, Y, produced by the P1R program; whereas, the P4F program produces a multiway frequency display which uses the P5D output statistics with the predictor values of the test sets to form a contingency table from which the verification scores and other statistics of interest can be

[8]The method was described in a private conversation and developed for this study based on a set of example programs provided by Lowe.

obtained. The procedure followed for this method is outlined below.

1) The P1R program performs a dimension reduction using linear regression. The regression equation which is derived from the predictor values of the training set by the P1R program is:

$$Y = C(0)+C(1)\bullet X(1)+C(2)\bullet X(2)+C(3)\bullet X(3), \qquad (4.1)$$

where C(0) is the intercept and C(1), C(2) and C(3) are the regression coefficients. This linear least squares fit produces a new variate, Y, which contains all the information of the three predictors. For each test set, the respective values of X(1), X(2) and X(3) predictors will yield a fresh value of Y.

2) The dimensionally-reduced Y variate is not a probability, but is used as an index or proxy in the discriminant procedures to be implemented next. The Y variate is grouped by the predictand categories and the P5D program is used to obtain the mean (μ) and variance (σ^2) statistics of the classification functions. Gaussian probability density functions are fit to these newly formed groups via the following equation:

$$L(m,n) = \exp\{-0.5 \bullet [(Y - \mu(m))/\sigma^2(m)]^2 - [(Y - \mu(n))/\sigma^2(n)]^2\}, \qquad (4.2)$$

where m,n = 1, 2, 3 forming six discriminant functions.

3) Then Bayes Law of Inverse Probability is used as a transform statement in the P4F program, using these six discriminant functions to discriminate the category of the predictand by choosing the one with the maximum probability value in Eqs.(4.3) through (4.5). These probabilities are given by:

$$P(1) = 1 / (1 + L(2,1) + L(3,1)) \qquad (4.3)$$

$$P(2) = 1 / (1 + L(1,2) + L(3,2)) \qquad (4.4)$$

$$P(3) = 1 / (1 + L(1,3) + L(2,3)) \qquad (4.5)$$

4. Discriminant Analysis Method

The Discriminant Analysis Method (DISC) was also proposed by Lowe[9] and uses Fisher's classical discriminant

[9] The DISC method was obtained in the same fashion as the DDRR method.

analysis design. The procedure consists of using the BMDP Statistical Software [University of California, 1983] programs - P7M and P4F. The P7M program performs a stepwise discriminant analysis of the input predictor distributions, while the P4F program is used, as with the DDRR method, to produce the multiway frequency table information. The procedure employed is.

1) Run the P7M program on the training set for the three predictors chosen by the PDM and obtain the classification functions C(i,j), where i is the index of the predictand category and j is the index of the predictor variable and the intercept.

2) Calculate from the set of classification functions the set of six discriminant functions needed to perform the discrimination by taking the differences between the elements to form the coefficients as seen in Eq.(4.6). Enter these values in the transform statement of the P4F program.

$$L(m,n)=\exp\{[C(m,1)-C(n,1)]\cdot X(1)+[C(m,2)-C(n,2)] \cdot X(2)+[C(m,3)-C(n,3)]\cdot X(3)+[C(m,4)-C(n,4)]\}, \quad (4.6)$$

where X(1), X(2) and X(3) are the variables to be filled by the three predictors of the test sets and m,n = 1, 2, 3.

3) Then these discriminant functions are used in Bayes Law of Inverse Probability to calculate the probabilities P(1), P(2) and P(3) of the predictor set belonging to categories 1, 2 and 3, respectively. Eqs.(4.3) through (4.5) are used here also to find the probabilities.

4) The P4F program using the discriminant functions and the test set predictors compiles a contingency table of the forecast versus observed categories of the predictand from which the verifying scores are calculated.

B. RESULTS

The successful simulation of the data, models and observers for this study is apparent from the verification scores[10] calculated. The perfect observer achieved the best scores for all the data sets (highest values for the A0 and TS1 scores, lowest values for the A1 score), while the bad observer had the worst scores; the good model did better

[10] See Appendix A for verification score definitions and comments.

than the bad model in all scores calculated; and the easy data outscored the hard data by a wide margin. These results are tabulated in Tables VI - VIII.

The PDM, DDRR and DISC methods appear to have remarkably similar values for the A0, A1 and TS1 verification scores. However, the MaxProb2 method has markedly different values. On the one hand, it scored better than the other three methods--higher values for A0 and TS1, lower values for A1--in the training set. This is probably due to the discrete way MaxProb2 pairs the predictand categories to the predictor intervals. On the other hand, it scored worse in the test sets. This gap between the training set scores and the test set scores occurs only with the MaxProb2 method. It appears evident that the MaxProb2 method will provide an extremely good fit to the training set data, if given enough intervals. However, if the least-squares-fit of the data of the testing set does not match that of the training set, then the verification scores show minimal skill. The other three methods have lesser differences between the training set and their test sets.

V. ANALYSIS OF VARIANCE (ANOVA) INTERCOMPARISONS OF THE MOS PREDICTION METHODS

A. INTRODUCTION TO ANOVA

The variations of one or more factors (prediction method, data set, model, observer) involved in this MOS study can be analyzed effectively by the technique of analysis of variance (ANOVA). This technique allows the variance of the measured variable (in this case, skill score) to be broken down into the portions caused by several factors, and interactions of those factors, whether varied singly or in combination, and a portion attributed to experimental error. ANOVA consists of:

1) A partitioning of the total sum of the squares of deviations of the skill score from the mean into two or more component sums of squares, each of which is associated with a particular factor or with experimental error, and

2) a parallel partitioning of the total number of degrees of freedom.

When certain variations of a factor are singled out for study because they are considered to be of more importance or interest, then ANOVA can be used as a comparison of the mean effects of those certain variations. Statistical tests (F tests) are made to determine whether the observed differences are probably real. If the differences are judged real, the main effects and interactions of the population may be estimated quite easily.

The ANOVA table utilized is described below.[11]

1) The first column lists the sources of variation and indicates which of the sources are being varied. For example, source ABC has three sources of variation with source D being held fixed.

[11] For further explanation of the ANOVA printouts, see Box, Hunter and Hunter (1978).

2) The second column lists the sums of the squares from all sources listed in column one.

3) The third column lists the number of degrees of freedom associated with the sources in column one and is calculated by subtracting one from the number of items being varied.

4) The fourth column lists the mean of the square deviations and is calculated by dividing the sums of squares by the degrees of freedom.

5) The fifth column lists the F test value observed for the sources in column one.

6) The sixth column lists the F test critical value which must be lower than the value in the fifth column if significance is to be shown.

7) The seventh column lists the Pvalue which is the actual probability that the variation observed is due to chance. Thus, a very low value would indicate that the observed variation is not due to chance.

B. THE INTERCOMPARISONS OF ALL FOUR METHODS USING ANOVA

Tables IX - XI contain the intercomparisons of the four MOS prediction methods using ANOVA, for the A0, A1 and TS1 scores, respectively. In these tables it can be seen that the variability source A (prediction method) is significant as compared to chance since the Fobs value is greater than the Fcrit value. Thus, there is a real difference between the methods that is more than just random. The large values evident for the Fobs terms for variance sources B, C and D (data, model and observer, respectively) are due to the differences simulated in their generation. Of interest is the fact that the variation due to the differences in the data sets is nearly twice that of the variations due to the differences in the model versions or the observer types. The negative values on the tables are due to the computer not properly handling very small values.

C. THE INTERCOMPARISONS OF THREE METHODS, MAXPROB2 REMOVED

The large Fobs term in the preceding tables for source A was due to differences between the methods. As seen in Chapter IV, Section B, the only method that appears to have

markedly different values is the MaxProb2 method. The results of an ANOVA comparison, when MaxProb2 is removed, can be seen in Tables XII - XIV, for the A0, A1 and TS1 scores. It is to be noted that the Fobs terms for source A in each is now less than the Fcrit terms, indicating that there is no significant difference between the three discriminant methods. The data variation is still nearly twice as large as the other two simulated variations.

D. THE INTERCOMPARISONS OF THREE METHODS, 2X2

Finally, a third set of ANOVA tables, this time a 2x2 comparison of the three remaining MOS prediction methods for each of the verification scores A0, A1 and TS1, can be seen in Tables XV - XXIII. This last comparation shows that there is no significant difference between either the PDM, DDRR or DISC methods, for any of the verification scores, regardless of how the methods are grouped.

VI. CONCLUSIONS AND RECOMMENDATIONS

A. CONCLUSIONS

First, the DDRR, DISC and PDM all have similar skill for all variations of model, observer and data when tested on the simulated data.

Second, the MaxProb2 method is not on the same performance level as the three other methods tested, and should be modified in future test studies with the recommendation by Preisendorfer that follows (see the final recommendation below).

Third, further co-evaluation of the three remaining MOS prediction methods should be continued with a goal of identifying statistically significant MOS predictive schemes for specific forecast problems.

Finally, the results from the Anova intercomparisons indicate that the variation due to the difference in the two data sets is the most important factor to be considered when seeking skillful MOS prediction schemes. Therefore, data collected must be carefully analyzed by objective methods (so as to identify those predictors whose values possess a greater separability when grouped according to predictand categories) in order for the resulting MOS forecast to be of potentially high operational worth.

B. RECOMMENDATIONS

First, the predictive skill of the MOS methods, when predictand categories have unequal frequencies, needs to be addressed. For example, the rare event (one which occurs very infrequently, i.e., less than 10% of the time) is of particular interest.

Second, the predictive skill of the MOS methods, when the predictor's class conditional populations (as determined by the predictand categories) have significantly unequal variances, also needs to be addressed.

Third, different types of parametric probability distributions, other than the normal distribution used to generate the data fields, need to be developed so as to ascertain the relative skill of the MOS predictive methods when confronted with significantly non-normal data. It may be important in the future to relax the spatially stationary feature in the simulated data equations in order to achieve this recommendation.

Fourth, predictand formulation methods other than the predominantly linear In-House Field method used in this study need to be investigated. The scientific literature contains many specific algorithms, which are non-linear, that connect real primary and real secondary fields (e.g., vapor pressure, humidity, wind, etc.), these can be profitably used in future simulation studies.

Fifth, provide a stronger foundation for future MOS studies by developing new techniques for the screening of predictors and predictor selection.

Sixth, the stochastic skill of the MOS prediction methods needs to be examined. In this study, a categorical forecasting procedure was used where the skill statistics are computed from a contingency table. However, an alternate forecasting procedure would use the actual predicted probabilities of belonging to a given predictand category. With such a probabilistic method, the stochastic skill is defined by the sharpness of the forecasted probabilities. For example, consider a two-category problem where one method predicts a 90% probability of belonging to category one and 10% to category two. Then this method is said to be sharper and of higher predictive skill than a method which

forecasts corresponding probabilities of 55% and 45% for the same problem.

Finally, the Maximum-Probability Method II strategy of choosing the category with maximum probability may not always be optimal. An alternate scheme would be to choose a category randomly, using its computed probability as a guide for the choice by a random number generator. Such a tactic would likely produce higher skill scores, overall, i.e., on average, than the MaxProb2 strategy. The basis for this can be demonstrated mathematically, using the Brier (1950) skill score, when the predictions are in terms of the probability of a category. However, it may also be described intuitively. Suppose, e.g., that the three probabilities of the predictand are 0.2, 0.5 and 0.3 for bad, marginal and good visibility for a certain realized predictor. The MaxProb2 strategy always directs the selection of the marginal category. If this strategy is followed many times, then the marginal category will be picked 100% of the time, and the low or high categories will not ever be picked. But the latter two occur 50% of the time, collectively. On the other hand, randomly choosing categories will allow the low and high categories, collectively, to be chosen 50% of the time. Clearly, by including the low and high categories in this way, a higher skill score would result. This new strategy[12] should be implemented henceforth in all further studies of the modified MaxProb2 method.

[12] This new strategy and its justification were proposed by Preisendorfer in a private conversation. Its implications have not yet been verified through experimentation.

APPENDIX A

STATISTICAL DEFINITIONS

1. VERIFICATION SCORES

See the table below for the following verification score definitions. Total=R + S + T + U + V + W + X + Y + Z

A) A0 - 'A naught' score - describes the probability of making a correct forecast given the total sample of observed events (also known as the Total Percent Correct score).

$$A0 = (R + V + Z) / Total$$

B) A1 - 'A one' score - describes the probability of a one category error which is made when a forecast is one category away from what was actually observed, i.e., category 2 forecast and either category 1, or 3 verified.

$$A1 = (S + U + W + Y) / Total$$

C) TS1 - 'Threat score' - describes the reduction of threat of being surprised by a category 1 event. In terms of set theory, it is the intersection of category 1 forecast value divided by the union of the observed and forecast category 1 values.

$$TS1 = R / (R + S + T + U + X)$$

Sample Contingency Table

		OBSERVED			
		1	2	3	
FORECAST	1	R	S	T	F1
	2	U	V	W	F2
	3	X	Y	Z	F3
		O1	O2	O3	TOTAL

2. PREDICTAND CATEGORIES

The equal variance definition of the predictand categories implies that the single-trial probability of success of a random forecaster (th: stochaster) who is forecasting these categories, is not the same as that for the equally populous definition of the predictand categories. The expected value of A0 by the stochaster for the presently constructed equal-variance categories is 0.5136, and for the 5% upper critical value it is 0.5956. The expected value of A1 for the equal-variance is 0.4352, and the 5% lower critical value is 0.3532.

These numbers are essential for an understanding of the verification scores in Tables VI (for A0) and VII (for A1). In particular, they tell us whether or not the observer, model and data sets have been simulated in a reasonable manner. For example, one would expect that the perfect observer, working with a good model and easy data, will obtain significantly high A0 scores and significantly low A1 scores for just about any reasonably competitive prediction scheme. This is borne out on perusal of Tables VI and VII. For instance, the perfect observer, using a good model and easy data yields, for the PDM method in test set 1, an A0 score of 0.801, far above the 5% upper critical value of 0.5956; moreover, the A1 score, in this case, is 0.199, far below the 5% lower critical value of 0.3532. As another instance for the PDM method in test set 1, a bad observer working with a bad model and hard data generates an A0 score of 0.478, which is below the 5% upper critical value of 0.5956, and indeed less than the average (expected) value of 0.5136. The A1 score, in this same case, is 0.389, which exceeds the 5% lower critical value of 0.3532.

Table VIII for TS1 cannot be as readily interpreted as the Tables for A0 and A1. This is because the average value of TS1 and its upper 5% critical value for the

stochaster are not readily derived for the equal-variance categories. They may, e.g., be worked out by Monte Carlo means for moderate sample sizes. For large sample sizes, asymptotic analytic estimates are possible, however, these will not be made here.

APPENDIX B

FIGURES

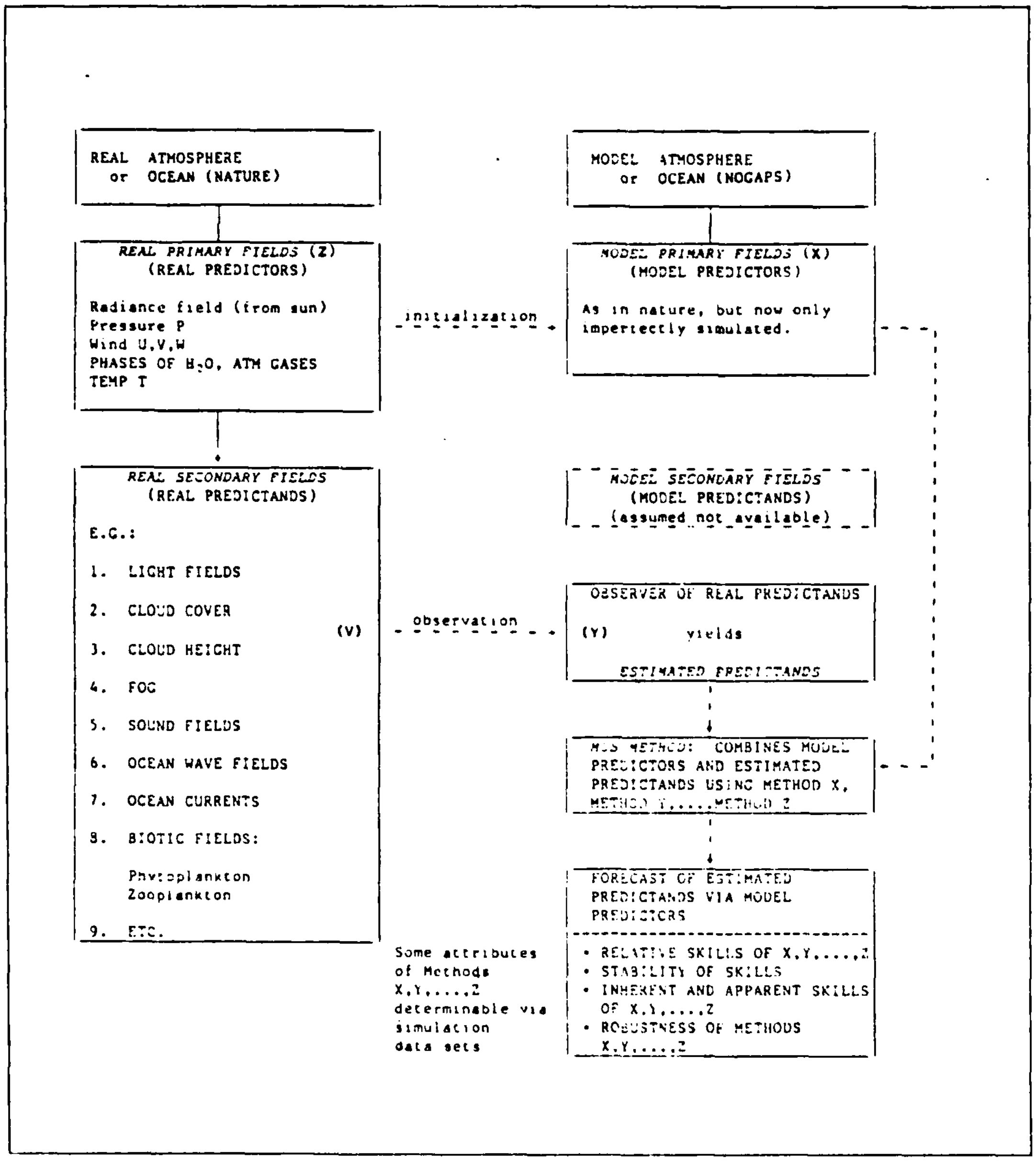


Fig. B.1 The Generalized Model Output Statistics Approach to Forecasting.

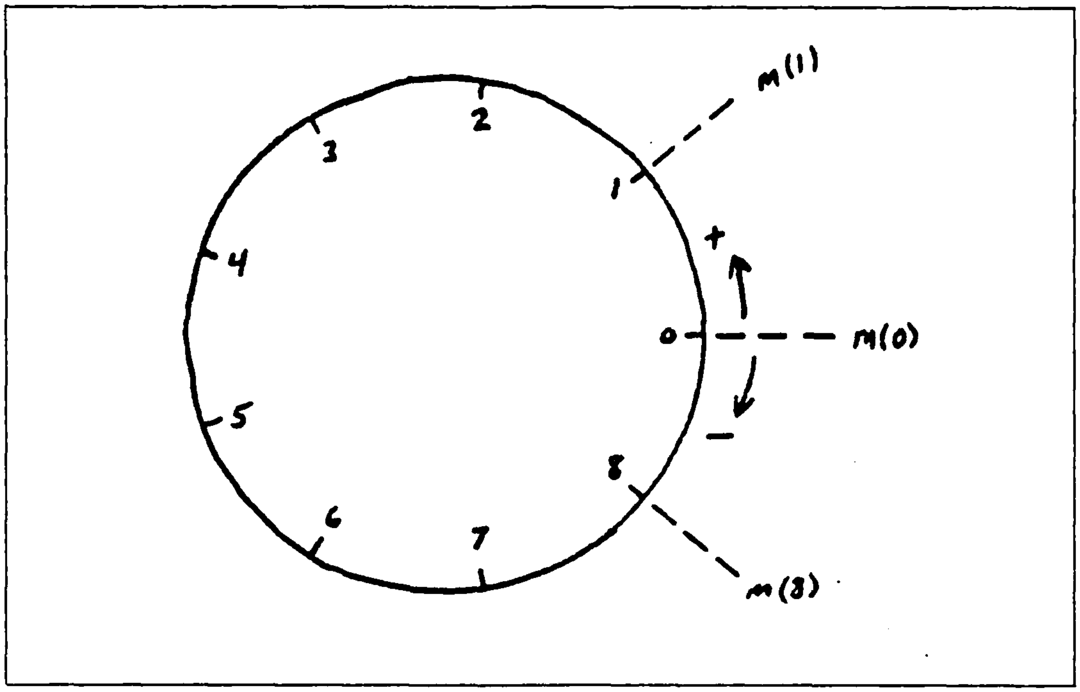


Fig. B.2   The Circle Modulo-p.

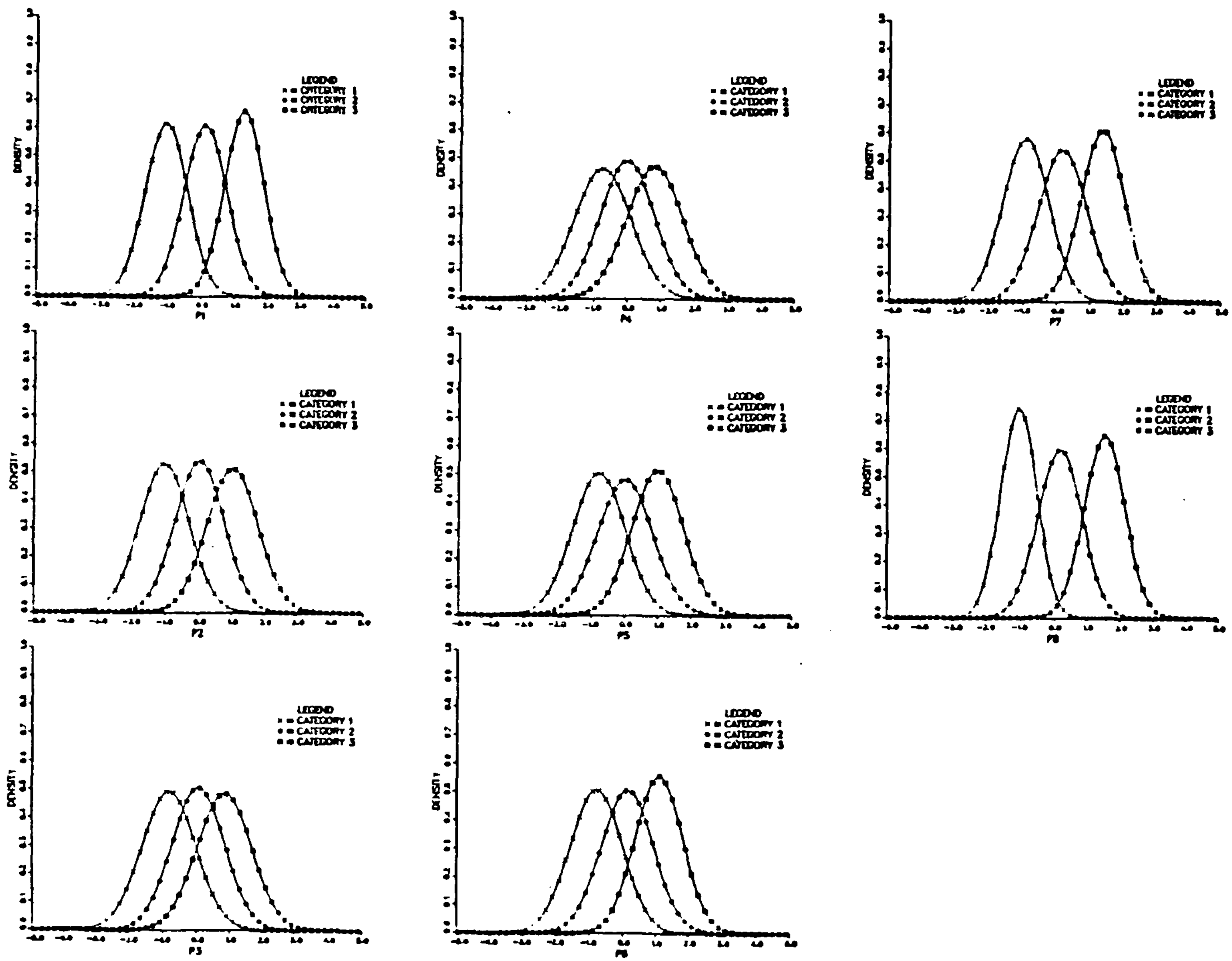


Fig. B.3 Correlations of the Eight Predictors of the Easy Data Set.

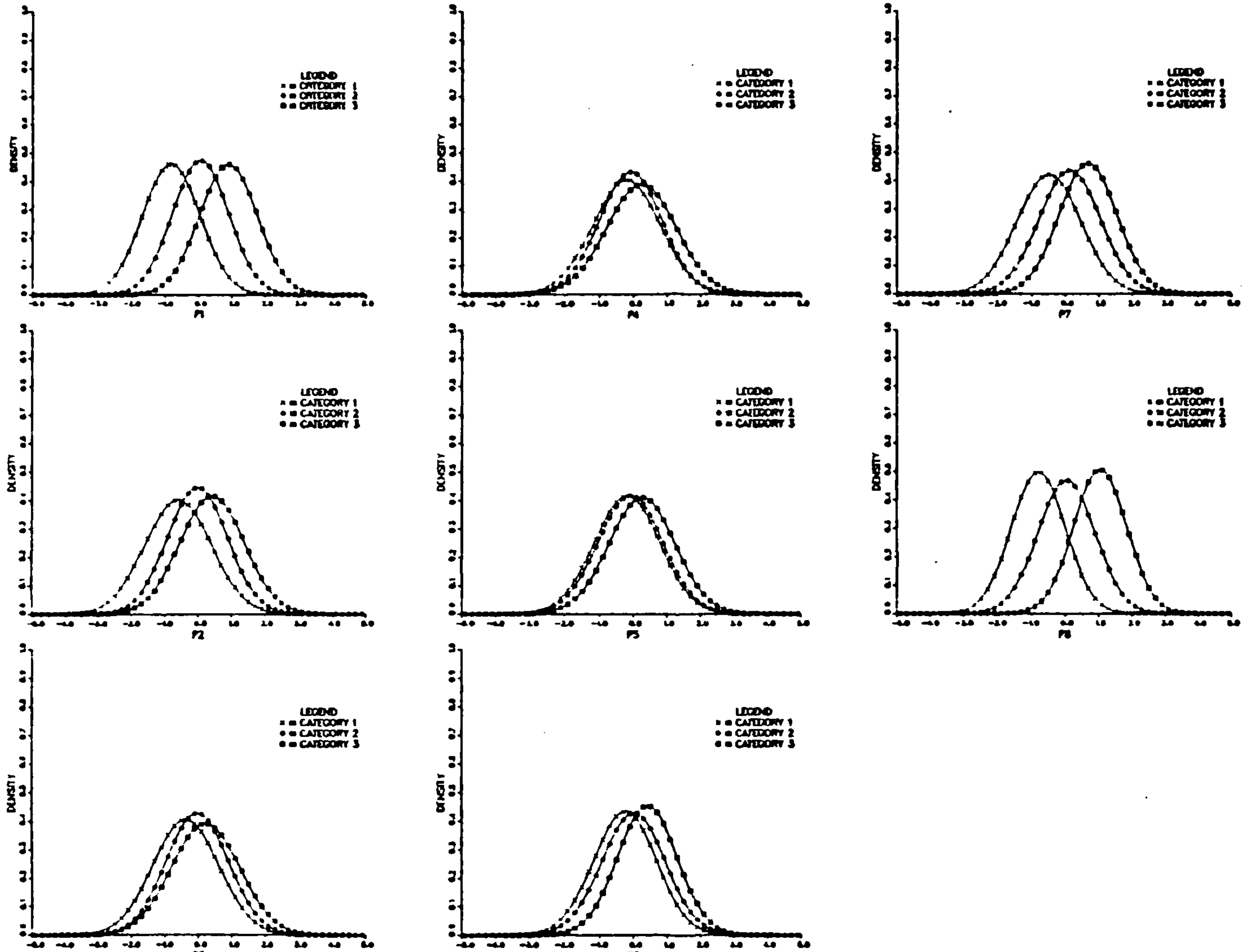


Fig. B.4 Correlations of the Eight Predictors of the Hard Data Set.

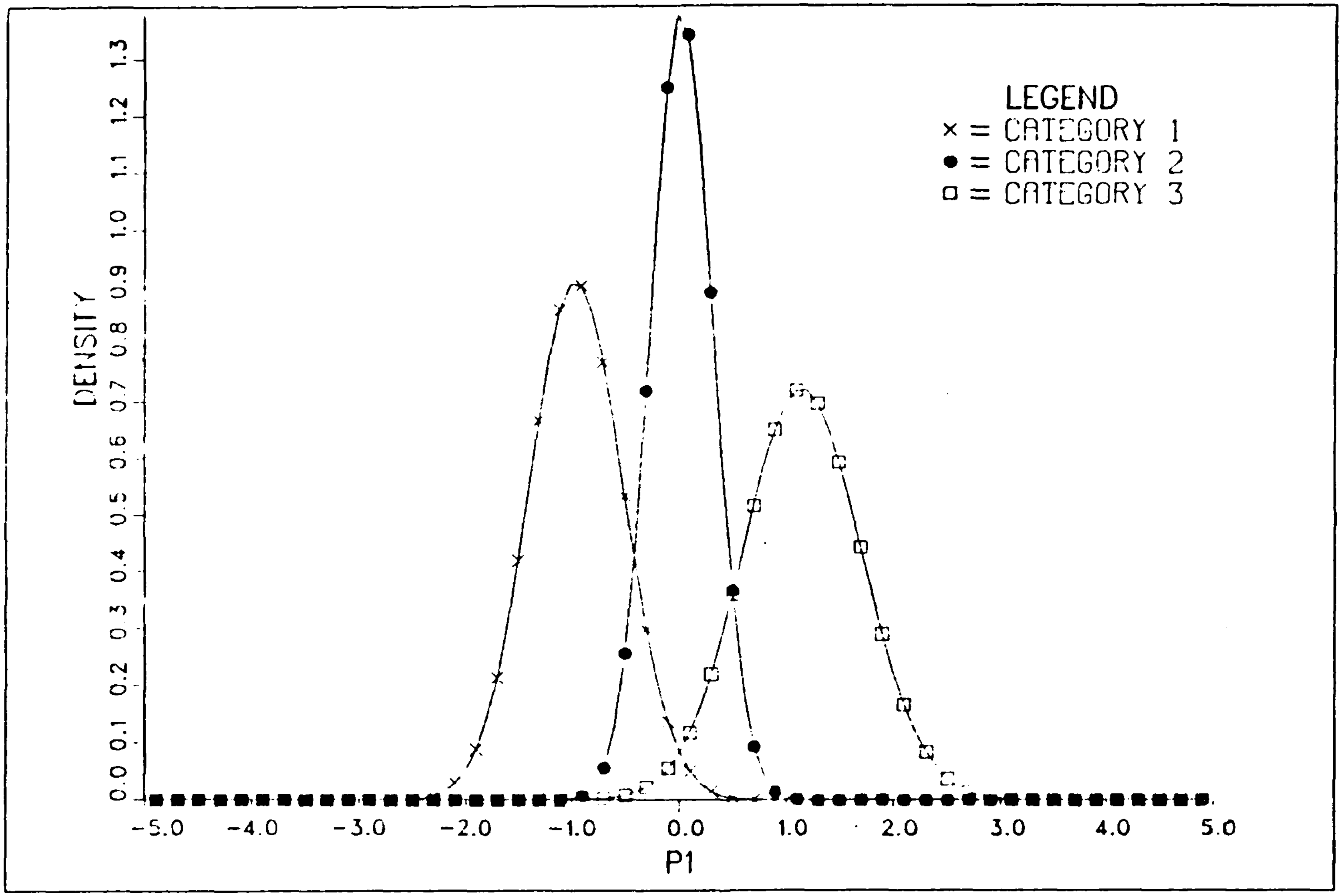


Fig. B.5 Equally Populous Interval (400 rows per Category).

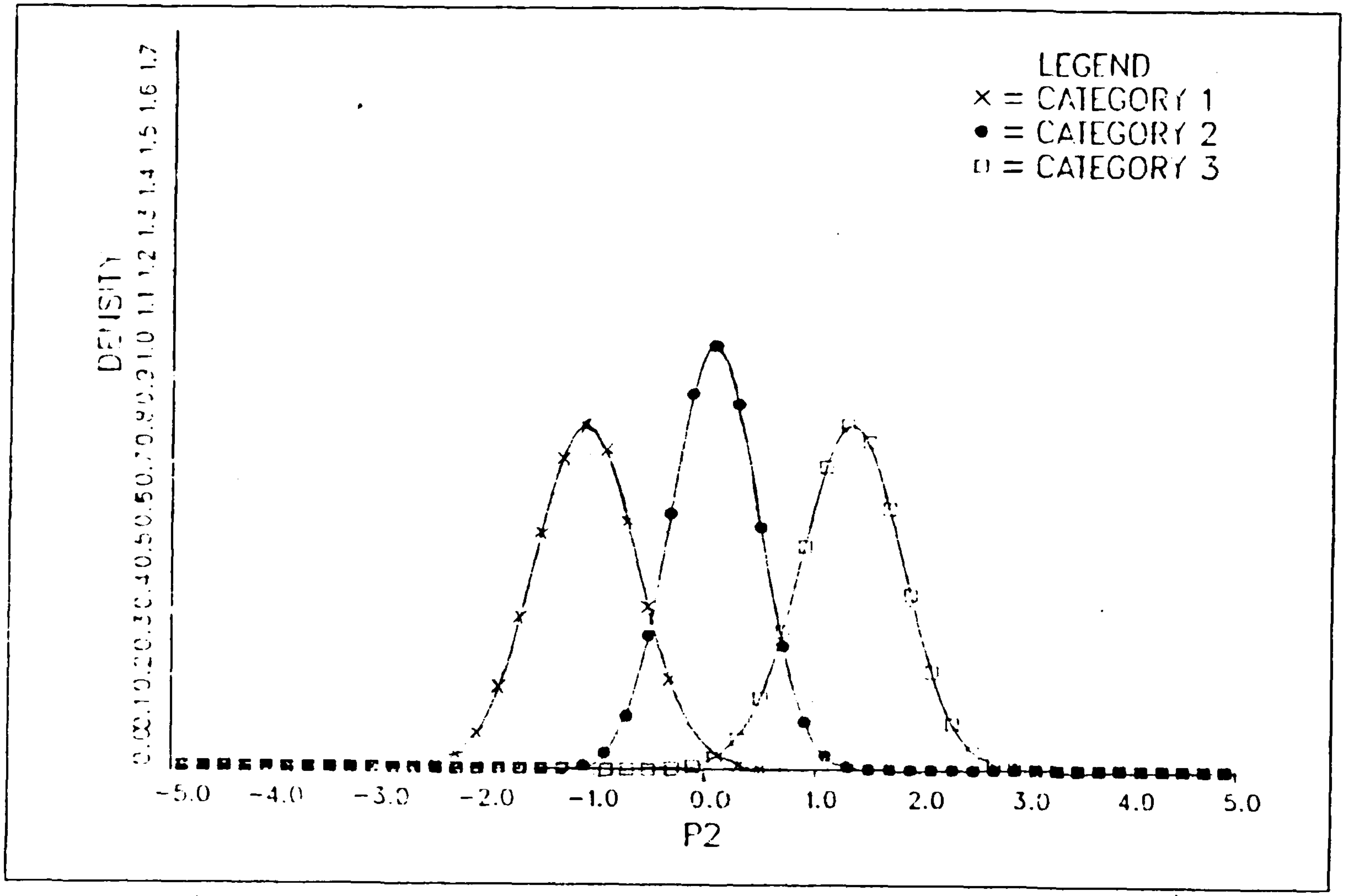


Fig. B.6 Equal Sigma Interval (1700 rows in Category 2).

APPENDIX C

TABLES

TABLE I

COVARIANCE MATRICES USED IN SIMULATING THE DATA SETS

EASY DATA SET

MATRIX M $\mu = 0.15$ $\kappa_a = 0.90$

(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1.000	0.861	0.741	0.638	0.549	0.472	0.407	0.350	0.301
0.861	1.000	0.861	0.741	0.638	0.549	0.472	0.407	0.350
0.741	0.861	1.000	0.861	0.741	0.638	0.549	0.472	0.407
0.638	0.741	0.861	1.000	0.861	0.741	0.638	0.549	0.472
0.549	0.638	0.741	0.861	1.000	0.861	0.741	0.638	0.549
0.472	0.549	0.638	0.741	0.861	1.000	0.861	0.741	0.638
0.407	0.472	0.549	0.638	0.741	0.861	1.000	0.861	0.741
0.350	0.407	0.472	0.549	0.638	0.741	0.861	1.000	0.861
0.301	0.350	0.407	0.472	0.549	0.638	0.741	0.861	1.000

MATRIX K

(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
0.900	0.775	0.667	0.574	0.494	0.425	0.366	0.315	0.271
0.775	0.900	0.775	0.667	0.574	0.494	0.425	0.366	0.315
0.667	0.775	0.900	0.775	0.667	0.574	0.494	0.425	0.366
0.574	0.667	0.775	0.900	0.775	0.667	0.574	0.494	0.425
0.494	0.574	0.667	0.775	0.900	0.775	0.667	0.574	0.494
0.425	0.494	0.574	0.667	0.775	0.900	0.775	0.667	0.574
0.366	0.425	0.494	0.574	0.667	0.775	0.900	0.775	0.667
0.315	0.366	0.425	0.494	0.574	0.667	0.775	0.900	0.775
0.271	0.315	0.366	0.425	0.494	0.574	0.667	0.775	0.900

HARD DATA SET

MATRIX M $\mu = 0.50$ $\kappa_a = 0.60$

(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1.000	0.606	0.368	0.223	0.135	0.082	0.050	0.030	0.018
0.606	1.000	0.606	0.368	0.223	0.135	0.082	0.050	0.030
0.368	0.606	1.000	0.606	0.368	0.223	0.135	0.082	0.050
0.223	0.368	0.606	1.000	0.606	0.368	0.223	0.135	0.082
0.135	0.223	0.368	0.606	1.000	0.606	0.368	0.223	0.135
0.082	0.135	0.223	0.368	0.606	1.000	0.606	0.368	0.223
0.050	0.082	0.135	0.223	0.368	0.606	1.000	0.606	0.368
0.030	0.050	0.082	0.135	0.223	0.368	0.606	1.000	0.606
0.018	0.030	0.050	0.082	0.135	0.223	0.368	0.606	1.000

MATRIX K

(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
0.600	0.364	0.221	0.134	0.081	0.049	0.030	0.018	0.011
0.364	0.600	0.364	0.221	0.134	0.081	0.049	0.030	0.018
0.221	0.364	0.600	0.364	0.221	0.134	0.081	0.049	0.030
0.134	0.221	0.364	0.600	0.364	0.221	0.134	0.081	0.049
0.081	0.134	0.221	0.364	0.600	0.364	0.221	0.134	0.081
0.049	0.081	0.134	0.221	0.364	0.600	0.364	0.221	0.134
0.030	0.049	0.081	0.134	0.221	0.364	0.600	0.364	0.221
0.018	0.030	0.049	0.081	0.134	0.221	0.364	0.600	0.364
0.011	0.018	0.030	0.049	0.081	0.134	0.221	0.364	0.600

TABLE II

COVARIANCE MATRICES AFTER MODIFICATION BY MODULAR ARITHMETIC

EASY DATA SET

MATRIX M		MATRIX K	
M(0) =	1.0000	K(0) =	0.9000
M(1) =	0.8607	K(1) =	0.7746
M(2) =	0.7408	K(2) =	0.6667
M(3) =	0.6376	K(3) =	0.5739
M(4) =	0.5488	K(4) =	0.4939
M(5) =	0.5488	K(5) =	0.4939
M(6) =	0.6376	K(6) =	0.5739
M(7) =	0.7408	K(7) =	0.6667
M(8) =	0.8607	K(8) =	0.7746

HARD DATA SET

MATRIX M		MATRIX K	
M(0) =	1.0000	K(0) =	0.6000
M(1) =	0.6065	K(1) =	0.3639
M(2) =	0.3679	K(2) =	0.2207
M(3) =	0.2231	K(3) =	0.1339
M(4) =	0.1353	K(4) =	0.0812
M(5) =	0.1353	K(5) =	0.0812
M(6) =	0.2231	K(6) =	0.1339
M(7) =	0.3679	K(7) =	0.2207
M(8) =	0.6065	K(8) =	0.3639

TABLE III

S_τ MATRICES USED TO DISTORT THE ORIGINAL SIGNAL

DISTORTION FOR THE GOOD MODEL

	1	2	3	4	5	6	7	8
1	0.95	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	0.95	0.00	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	0.95	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.00	0.95	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.95	0.00	0.00	0.00
6	0.00	0.00	0.00	0.00	0.00	0.95	0.00	0.00
7	0.00	0.00	0.00	0.00	0.00	0.00	0.95	0.00
8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.95

DISTORTION FOR THE BAD MODEL

	1	2	3	4	5	6	7	8
1	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	0.50	0.00	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	0.50	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.00	0.50	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.50	0.00	0.00	0.00
6	0.00	0.00	0.00	0.00	0.00	0.50	0.00	0.00
7	0.00	0.00	0.00	0.00	0.00	0.00	0.50	0.00
8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.50

TABLE IV

OBSERVER ERROR TABLE (OET)

e_{ij}	1	2	3	
Estimated category i	e_{11}	e_{12}	e_{13}	1
	e_{21}	e_{22}	e_{23}	2
	e_{31}	e_{32}	e_{33}	3

Actual category j

TABLE V

SKILLS OF THE THREE OBSERVER TYPES

PERFECT OBSERVER			GOOD OBSERVER			BAD OBSERVER		
1.00	0.00	0.00	0.90	0.10	0.00	0.72	0.18	0.10
0.00	1.00	0.00	0.10	0.80	0.10	0.18	0.64	0.18
0.00	0.00	1.00	0.00	0.10	0.90	0.10	0.18	0.72

TABLE VI

VERIFICATION SCORES FOR A0

GOOD MODEL - EASY DATA

OBSERVER	TRAINING SET				TEST SET 1				TEST SET 2				TEST SET 3			
METHODS	PDM	DISC	DDRR	MAXP	PDM	DISC	DDRR	MAXP	PDM	DISC	DDRR	MAXP	PDM	DISC	DDRR	MAXP
PERFECT	.819	.817	.813	.883	.801	.799	.795	.731	.810	.813	.816	.768	.777	.780	.775	.732
GOOD	.728	.733	.729	.828	.708	.710	.708	.660	.726	.725	.730	.665	.701	.692	.697	.658
BAD	.593	.587	.584	.682	.572	.582	.582	.543	.572	.573	.579	.538	.542	.564	.565	.533

BAD MODEL - EASY DATA

OBSERVER	TRAINING SET				TEST SET 1				TEST SET 2				TEST SET 3			
METHODS	PDM	DISC	DDRR	MAXP	PDM	DISC	DDRR	MAXP	PDM	DISC	DDRR	MAXP	PDM	DISC	DDRR	MAXP
PERFECT	.669	.668	.669	.856	.657	.659	.659	.573	.705	.703	.708	.585	.678	.681	.682	.590
GOOD	.632	.630	.632	.796	.597	.602	.602	.549	.648	.660	.654	.565	.633	.633	.632	.553
BAD	.542	.527	.539	.838	.527	.521	.533	.453	.565	.563	.549	.447	.556	.552	.552	.463

GOOD MODEL - *HARD DATA*

OBSERVER	TRAINING SET				TEST SET 1				TEST SET 2				TEST SET 3			
METHODS	PDM	DISC	DDRR	MAXP	PDM	DISC	DDRR	MAXP	PDM	DISC	DDRR	MAXP	PDM	DISC	DDRR	MAXP
PERFECT	.632	.642	.640	.862	.640	.639	.639	.546	.643	.642	.653	.577	.634	.634	.642	.553
GOOD	.598	.601	.597	.897	.590	.596	.588	.491	.599	.607	.611	.533	.580	.588	.587	.492
BAD	.517	.520	.520	.932	.500	.493	.517	.436	.487	.49[illegible]	.520	.413	.497	.503	.510	.423

BAD MODEL - HARD DATA

OBSERVER	TRAINING SET				TEST SET 1				TEST SET 2				TEST SET 3			
METHODS	PDM	DISC	DDRR	MAXP	PDM	DISC	DDRR	MAXP	PDM	DISC	DDRR	MAXP	PDM	DISC	DDRR	MAXP
PERFECT	.543	.542	.535	.917	.543	.548	.549	.461	.537	.537	.548	.467	.532	.529	.535	.478
GOOD	.520	.521	.518	.830	.522	.523	.524	.444	.518	.521	.517	.427	.510	.508	.512	.423
BAD	.467	.476	.472	.945	.478	.464	.469	.397	.472	.467	.458	.370	.463	.455	.457	.409

TABLE VII

VERIFICATION SCORES FOR A1

GOOD MODEL - EASY DATA

OBSERVER	TRAINING SET				TEST SET 1				TEST SET 2				TEST SET 3			
	PDM	DISC	DDRR	MAXP	PDM	DISC	DDRR	MAXP	PDM	DISC	DDRR	MAXP	PDM	DISC	DDRR	MAXP
PERFECT	.180	.182	.187	.115	.199	.201	.205	.262	.190	.187	.184	.229	.222	.220	.225	.268
GOOD	.257	.252	.252	.154	.277	.278	.277	.312	.260	.261	.255	.314	.290	.297	.292	.309
BAD	.328	.311	.319	.240	.353	.313	.319	.347	.352	.322	.320	.355	.385	.343	.344	.350
	METHODS				METHODS				METHODS				METHODS			

BAD MODEL - EASY DATA

OBSERVER	TRAINING SET				TEST SET 1				TEST SET 2				TEST SET 3			
	PDM	DISC	DDRR	MAXP	PDM	DISC	DDRR	MAXP	PDM	DISC	DDRR	MAXP	PDM	DISC	DDRR	MAXP
PERFECT	.322	.323	.321	.130	.332	.330	.330	.374	.291	.292	.287	.378	.311	.310	.307	.365
GOOD	.339	.342	.340	.172	.377	.373	.372	.379	.329	.320	.324	.380	.337	.342	.340	.386
BAD	.363	.367	.366	.118	.370	.367	.369	.398	.342	.329	.356	.401	.342	.335	.341	.392
	METHODS				METHODS				METHODS				METHODS			

GOOD MODEL - HARD DATA

OBSERVER	TRAINING SET				TEST SET 1				TEST SET 2				TEST SET 3			
	PDM	DISC	DDRR	MAXP	PDM	DISC	DDRR	MAXP	PDM	DISC	DDRR	MAXP	PDM	DISC	DDRR	MAXP
PERFECT	.345	.337	.338	.122	.335	.335	.331	.392	.337	.343	.330	.369	.342	.342	.336	.384
GOOD	.365	.366	.364	.082	.360	.359	.362	.428	.359	.358	.352	.392	.365	.366	.357	.421
BAD	.377	.364	.377	.052	.383	.377	.353	.387	.394	.378	.363	.430	.371	.356	.364	.420
	METHODS				METHODS				METHODS				METHODS			

BAD MODEL - HARD DATA

OBSERVER	TRAINING SET				TEST SET 1				TEST SET 2				TEST SET 3			
	PDM	DISC	DDRR	MAXP	PDM	DISC	DDRR	MAXP	PDM	DISC	DDRR	MAXP	PDM	DISC	DDRR	MAXP
PERFECT	.381	.373	.387	.064	.378	.372	.381	.413	.392	.382	.382	.402	.398	.397	.401	.381
GOOD	.374	.374	.376	.129	.381	.387	.377	.427	.384	.387	.385	.425	.391	.400	.387	.440
BAD	.404	.362	.394	.037	.389	.377	.400	.436	.395	.370	.410	.434	.397	.377	.403	.392
	METHODS				METHODS				METHODS				METHODS			

TABLE VIII

VERIFICATION SCORES FOR TS1

GOOD MODEL - EASY DATA

OBSERVER	TRAINING SET				TEST SET 1				TEST SET 2				TEST SET 3			
METHODS	PDM	DISC	DDRR	MAXP	PDM	DISC	DDRR	MAXP	PDM	DISC	DDRR	MAXP	PDM	DISC	DDRR	MAXP
PERFECT	.770	.773	.763	.835	.772	.771	.764	.707	.771	.768	.778	.730	.766	.767	.765	.740
GOOD	.681	.680	.683	.773	.649	.661	.653	.620	.663	.661	.664	.607	.665	.670	.661	.623
BAD	.504	.507	.508	.586	.485	.495	.487	.440	.479	.498	.494	.447	.459	.489	.485	.465

BAD MODEL - EASY DATA

OBSERVER	TRAINING SET				TEST SET 1				TEST SET 2				TEST SET 3			
METHODS	PDM	DISC	DDRR	MAXP	PDM	DISC	DDRR	MAXP	PDM	DISC	DDRR	MAXP	PDM	DISC	DDRR	MAXP
PERFECT	.590	.587	.589	.793	.581	.580	.583	.484	.663	.658	.666	.513	.655	.663	.655	.539
GOOD	.541	.539	.545	.708	.510	.517	.520	.454	.584	.595	.592	.499	.583	.587	.582	.483
BAD	.459	.453	.453	.768	.426	.435	.428	.365	.485	.478	.465	.366	.474	.470	.462	.388

GOOD MODEL - HARD DATA

OBSERVER	TRAINING SET				TEST SET 1				TEST SET 2				TEST SET 3			
METHODS	PDM	DISC	DDRR	MAXP	PDM	DISC	DDRR	MAXP	PDM	DISC	DDRR	MAXP	PDM	DISC	DDRR	MAXP
PERFECT	.553	.567	.558	.788	.599	.594	.586	.494	.599	.608	.609	.496	.596	.589	.594	.492
GOOD	.525	.523	.526	.826	.531	.529	.526	.431	.549	.555	.565	.465	.515	.519	.520	.440
BAD	.422	.428	.415	.885	.405	.407	.423	.335	.414	.417	.438	.337	.423	.429	.428	.363

BAD MODEL - HARD DATA

OBSERVER	TRAINING SET				TEST SET 1				TEST SET 2				TEST SET 3			
METHODS	PDM	DISC	DDRR	MAXP	PDM	DISC	DDRR	MAXP	PDM	DISC	DDRR	MAXP	PDM	DISC	DDRR	MAXP
PERFECT	.450	.452	.449	.854	.451	.468	.463	.357	.458	.463	.476	.382	.453	.455	.462	.391
GOOD	.427	.430	.429	.741	.430	.430	.434	.344	.441	.444	.446	.329	.433	.430	.432	.351
BAD	.361	.382	.369	.900	.385	.381	.372	.307	.380	.390	.381	.273	.371	.372	.365	.316

TABLE IX

ANOVA TABLE FOR A0 - 4 METHODS COMPARED

SOURCE	SUMSQ	DF	MS	FOBS	FCRIT	PVALUE
A	0.1451	3	0.0484	230.47	2.682	0.000000
B	0.4838	1	0.4838	2305.96	3.857	0.000000
C	0.2145	1	0.2145	1022.55	3.857	0.000000
D	0.4955	2	0.2478	1180.87	3.043	0.000000
AB	0.0014	3	0.0005	2.30	2.682	0.077623
AC	0.0027	3	0.0009	4.24	2.682	0.004303
AD	0.0003	6	0.0000	0.21	2.212	0.965467
BC	0.0020	1	0.0020	9.38	3.857	0.000764
BD	0.0306	2	0.0153	72.91	3.043	0.000000
CD	0.0344	2	0.0172	82.00	3.043	0.000000
ABC	0.0037	3	0.0012	5.82	2.682	0.000384
ABD	0.0001	6	0.0000	0.10	2.212	1.019332
ACD	0.0003	6	0.0000	0.23	2.212	0.954907
BCD	0.0003	2	0.0002	0.76	3.043	0.495613
ABCD	0.0016	6	0.0003	1.27	2.212	0.287966
ERROR	0.0201	96				
TOTAL	1.4364	143				

CO 48.6228 MSE 0.0002

ANOVA TABLE WITH 4 FACTORS:
A = 4 MOS METHODS
B = 2 DATA SETS
C = 2 MODEL VERSIONS
D = 3 OBSERVER TYPES

TABLE X

ANOVA TABLE FOR A1 - 4 METHODS COMPARED

SOURCE	SUMSQ	DF	MS	FOBS	FCRIT	PVALUE
A	0.0422	3	0.0141	66.15	2.682	0.000000
B	0.1626	1	0.1626	765.34	3.857	0.000000
C	0.0800	1	0.0800	376.60	3.857	0.000000
D	0.0645	2	0.0323	151.86	3.043	0.000000
AB	-0.0002	3	-0.0001	-0.24	2.682	0.852105
AC	0.0005	3	0.0002	0.84	2.682	0.500000
AD	0.0016	6	0.0003	1.29	2.212	0.278279
BC	0.0135	1	0.0135	63.47	3.857	0.000000
BD	0.0216	2	0.0108	50.87	3.043	0.000000
CD	0.0195	2	0.0098	45.92	3.043	0.000000
ABC	0.0022	3	0.0007	3.40	2.682	0.015532
ABD	0.0025	6	0.0004	1.95	2.212	0.077145
ACD	0.0021	6	0.0004	1.66	2.212	0.139072
BCD	0.0064	2	0.0032	15.08	3.043	0.000000
ABCD	0.0011	6	0.0002	0.86	2.212	0.542469
ERROR	0.0204	96				
TOTAL	0.4407	143				

CO 17.4203 MSE 0.0002

ANOVA TABLE WITH 4 FACTORS:
A = 4 MOS METHODS
B = 2 DATA SETS
C = 2 MODEL VERSIONS
D = 3 OBSERVER TYPES

TABLE XI

ANOVA TABLE FOR TS1 - 4 METHODS COMPARED

SOURCE	SUMSQ	DF	MS	FOBS	FCRIT	PVALUE
A	0.1663	3	0.0554	138.33	2.682	0.000000
B	0.5971	1	0.5971	1490.11	3.857	0.000000
C	0.3342	1	0.3342	834.07	3.857	0.000000
D	0.7522	2	0.3761	938.60	3.043	0.000000
AB	0.0024	3	0.0008	2.03	2.682	0.113553
AC	0.0048	3	0.0016	3.99	2.682	0.006372
AD	0.0016	6	0.0003	0.66	2.212	0.691236
BC	0.0009	1	0.0009	2.32	3.857	0.127956
BD	0.0583	2	0.0291	72.71	3.043	0.000000
CD	0.0627	2	0.0313	78.24	3.043	0.000000
ABC	0.0056	3	0.0019	4.70	2.682	0.002144
ABD	-0.0001	6	0.0000	-0.05	2.212	1.064746
ACD	0.0004	6	0.0001	0.15	2.212	0.989171
BCD	0.0005	2	0.0003	0.63	3.043	0.559566
ABCD	0.0023	6	0.0004	0.95	2.212	0.478893
ERROR	0.0385	96				
TOTAL	2.0277	143				

CO 37.9683 MSE 0.0004

ANOVA TABLE WITH 4 FACTORS:
A = 4 MOS METHODS
B = 2 DATA SETS
C = 2 MODEL VERSIONS
D = 3 OBSERVER TYPES

TABLE XII

ANOVA TABLE FOR AO - 3 METHODS COMPARED

SOURCE	SUMSQ	DF	MS	FOBS	FCRIT	PVALUE
A	0.0007	2	0.0003	1.50	3.078	0.244744
B	0.3424	1	0.3424	1560.90	3.890	0.000000
C	0.1423	1	0.1423	648.83	3.890	0.000000
D	0.3817	2	0.1908	870.02	3.078	0.000000
AB	0.0002	2	0.0001	0.45	3.078	0.651507
AC	0.0003	2	0.0002	0.70	3.078	0.527535
AD	-0.0001	4	0.0000	-0.10	2.503	0.965831
BC	0.0002	1	0.0002	0.77	3.890	0.408447
BD	0.0245	2	0.0122	55.79	3.078	0.000000
CD	0.0276	2	0.0138	62.85	3.078	0.000000
ABC	-0.0004	2	-0.0002	-0.90	3.078	0.434343
ABD	-0.0002	4	-0.0001	-0.28	2.503	0.879029
ACD	0.0003	4	0.0001	0.37	2.503	0.824636
BCD	0.0005	2	0.0002	1.08	3.078	0.369748
ABCD	0.0008	4	0.0002	0.96	2.503	0.456994
ERROR	0.0158	72				
TOTAL	0.9364	107				

CO 38.7965 MSE 0.0002

ANOVA TABLE WITH 4 FACTORS:
A = 3 MOS METHODS
B = 2 DATA SETS
C = 2 MODEL VERSIONS
D = 3 OBSERVER TYPES

TABLE XIII
ANOVA TABLE FOR A1 - 3 METHODS COMPARED

SOURCE	SUMSQ	DF	MS	FOBS	FCRIT	PVALUE
A	0.0011	2	0.0006	2.61	3.078	0.075201
B	0.1244	1	0.1244	580.57	3.890	0.000000
C	0.0607	1	0.0607	283.38	3.890	0.000000
D	0.0525	2	0.0262	122.51	3.078	0.000000
AB	-0.0002	2	-0.0001	-0.45	3.078	0.652335
AC	0.0005	2	0.0003	1.21	3.078	0.325188
AD	0.0012	4	0.0003	1.41	2.503	0.248677
BC	0.0059	1	0.0059	27.74	3.890	0.000000
BD	0.0214	2	0.0107	50.04	3.078	0.000000
CD	0.0175	2	0.0088	40.87	3.078	0.000000
ABC	0.0004	2	0.0002	0.88	3.078	0.446352
ABD	0.0004	4	0.0001	0.45	2.503	0.770252
ACD	0.0016	4	0.0004	1.87	2.503	0.124488
BCD	0.0055	2	0.0027	12.74	3.078	0.000002
ABCD	0.0003	4	0.0001	0.40	2.503	0.803020
ERROR	0.0154	72				
TOTAL	0.3088	107				

CO 12.3448 MSE 0.0002

ANOVA TABLE WITH 4 FACTORS:
A = 3 MOS METHODS
B = 2 DATA SETS
C = 2 MODEL VERSIONS
D = 3 OBSERVER TYPES

TABLE XIV

ANOVA TABLE FOR TS1 - 3 METHODS COMPARED

SOURCE	SUMSQ	DF	MS	FOBS	FCRIT	PVALUE
A	0.0007	2	0.0004	0.80	3.078	0.480915
B	0.4164	1	0.4164	946.79	3.890	0.000000
C	0.2175	1	0.2175	494.53	3.890	0.000000
D	0.5954	2	0.2977	676.99	3.078	0.000000
AB	0.0001	2	0.0001	0.14	3.078	0.838436
AC	0.0001	2	0.0000	0.09	3.078	0.873986
AD	-0.0002	4	0.0000	-0.09	2.503	0.974577
BC	0.0004	1	0.0004	0.80	3.890	0.398383
BD	0.0446	2	0.0223	50.66	3.078	0.000000
CD	0.0463	2	0.0231	52.64	3.078	0.000000
ABC	-0.0003	2	-0.0001	-0.33	3.078	0.722544
ABD	-0.0001	4	0.0000	-0.04	2.503	0.999907
ACD	0.0006	4	0.0002	0.36	2.503	0.830389
BCD	0.0000	2	0.0000	0.02	3.078	0.931326
ABCD	0.0007	4	0.0002	0.41	2.503	0.798596
ERROR	0.0317	72				
TOTAL	1.3538	107				

CO 30.6874 MSE 0.0004

ANOVA TABLE WITH 4 FACTORS:
A = 3 MOS METHODS
B = 2 DATA SETS
C = 2 MODEL VERSIONS
D = 3 OBSERVER TYPES

TABLE XV

ANOVA TABLE FOR AO - PDM VS DISC

SOURCE	SUMSQ	DF	MS	FOBS	FCRIT	PVALUE
A	0.0002	1	0.0002	0.95	3.958	0.356832
B	0.2333	1	0.2333	1036.68	3.958	0.000000
C	0.0901	1	0.0901	400.14	3.958	0.000000
D	0.2572	2	0.1286	571.42	3.149	0.000000
AB	0.0000	1	0.0000	0.07	3.958	0.736694
AC	0.0000	1	0.0000	0.07	3.958	0.736694
AD	-0.0002	2	-0.0001	-0.41	3.149	0.679774
BC	0.0000	1	0.0000	0.14	3.958	0.684829
BD	0.0162	2	0.0081	35.90	3.149	0.000000
CD	0.0216	2	0.0108	48.03	3.149	0.000000
ABC	-0.0002	1	-0.0002	-0.68	3.958	0.436663
ABD	0.0001	2	0.0000	0.17	3.149	0.818564
ACD	0.0002	2	0.0001	0.44	3.149	0.658203
BCD	0.0005	2	0.0002	1.02	3.149	0.394105
ABCD	0.0002	2	0.0001	0.54	3.149	0.604052
ERROR	0.0108	48				
TOTAL	0.6301	71				

CO 25.7722 MSE 0.0002

ANOVA TABLE WITH 4 FACTORS:
A = 2 METHODS (PDM VS DISC)
B = 2 DATA SETS
C = 2 MODEL VERSIONS
D = 3 OBSERVER TYPES

TABLE XVI

ANOVA TABLE FOR A1 - PDM VS DISC

SOURCE	SUMSQ	DF	MS	FOBS	FCRIT	PVALUE
A	0.0008	1	0.0008	3.80	3.958	0.046882
B	0.0825	1	0.0825	377.69	3.958	0.000000
C	0.0345	1	0.0345	157.92	3.958	0.000000
D	0.0349	2	0.0175	79.94	3.149	0.000000
AB	-0.0001	1	-0.0001	-0.35	3.958	0.569152
AC	-0.0001	1	-0.0001	-0.37	3.958	0.556986
AD	0.0014	2	0.0007	3.16	3.149	0.044791
BC	0.0046	1	0.0046	21.25	3.958	0.000003
BD	0.0154	2	0.0077	35.23	3.149	0.000000
CD	0.0158	2	0.0079	36.18	3.149	0.000000
ABC	0.0002	1	0.0002	1.13	3.958	0.312872
ABD	0.0002	2	0.0001	0.37	3.149	0.699473
ACD	0.0003	2	0.0002	0.69	3.149	0.530105
BCD	0.0035	2	0.0018	8.02	3.149	0.000363
ABCD	0.0003	2	0.0002	0.74	3.149	0.507911
ERROR	0.0105	48				
TOTAL	0.2047	71				

CO 8.2640 MSE 0.0002

ANOVA TABLE WITH 4 FACTORS:
A = 2 METHODS (PDM VS DISC)
B = 2 DATA SETS
C = 2 MODEL VERSIONS
D = 3 OBSERVER TYPES

TABLE XVII

ANOVA TABLE FOR TS1 - PDM VS DISC

SOURCE	SUMSQ	DF	MS	FOBS	FCRIT	PVALUE
A	0.0005	1	0.0005	1.01	3.958	0.340231
B	0.2840	1	0.2840	628.63	3.958	0.000000
C	0.1414	1	0.1414	312.96	3.958	0.000000
D	0.3942	2	0.1971	436.34	3.149	0.000000
AB	0.0000	1	0.0000	-0.10	3.958	0.709762
AC	-0.0001	1	-0.0001	-0.14	3.958	0.685636
AD	-0.0001	2	-0.0001	-0.14	3.149	0.842322
BC	0.0000	1	0.0000	0.03	3.958	0.770692
BD	0.0290	2	0.0145	32.11	3.149	0.000000
CD	0.0358	2	0.0179	39.67	3.149	0.000000
ABC	0.0000	1	0.0000	0.03	3.958	0.770692
ABD	0.0001	2	0.0001	0.15	3.149	0.829864
ACD	0.0003	2	0.0001	0.29	3.149	0.746003
BCD	0.0003	2	0.0001	0.29	3.149	0.746003
ABCD	0.0002	2	0.0001	0.19	3.149	0.808177
ERROR	0.0217	48				
TOTAL	0.9072	71				

CO 20.4221 MSE 0.0005

ANOVA TABLE WITH 4 FACTORS:
A = 2 METHODS (PDM VS DISC)
B = 2 DATA SETS
C = 2 MODEL VERSIONS
D = 3 OBSERVER TYPES

TABLE XVIII

ANOVA TABLE FOR AO - PDM VS DDRR

SOURCE	SUMSQ	DF	MS	FOBS	FCRIT	PVALUE
A	0.0004	1	0.0004	1.90	3.958	0.178464
B	0.2237	1	0.2237	1031.79	3.958	0.000000
C	0.0950	1	0.0950	438.05	3.958	0.000000
D	0.2521	2	0.1260	581.31	3.149	0.000000
AB	0.0001	1	0.0001	0.49	3.958	0.504978
AC	0.0003	1	0.0003	1.20	3.958	0.297287
AD	-0.0002	2	-0.0001	-0.35	3.149	0.710845
BC	0.0000	1	0.0000	0.14	3.958	0.681342
BD	0.0169	2	0.0085	38.99	3.149	0.000000
CD	0.0182	2	0.0091	41.91	3.149	0.000000
ABC	-0.0002	1	-0.0002	-0.84	3.958	0.383821
ABD	-0.0001	2	0.0000	-0.18	3.149	0.816188
ACD	0.0006	2	0.0003	1.48	3.149	0.252559
BCD	0.0008	2	0.0004	1.76	3.149	0.190655
ABCD	0.0003	2	0.0001	0.63	3.149	0.558304
ERROR	0.0104	48				
TOTAL	0.6183	71				

CO 25.8884 MSE 0.0002

ANOVA TABLE WITH 4 FACTORS:
A = 2 METHODS (PDM VS DDRR)
B = 2 DATA SETS
C = 2 MODEL VERSIONS
D = 3 OBSERVER TYPES

TABLE XIX

ANOVA TABLE FOR A1 - PDM VS DDRR

SOURCE	SUMSQ	DF	MS	FOBS	FCRIT	PVALUE
A	0.0006	1	0.0006	2.92	3.958	0.086440
B	0.0814	1	0.0814	392.98	3.958	0.000000
C	0.0424	1	0.0424	204.75	3.958	0.000000
D	0.0404	2	0.0202	97.63	3.149	0.000000
AB	-0.0001	1	-0.0001	-0.45	3.958	0.523927
AC	0.0006	1	0.0006	2.74	3.958	0.097799
AD	0.0002	2	0.0001	0.43	3.149	0.662482
BC	0.0030	1	0.0030	14.35	3.958	0.000073
BD	0.0146	2	0.0073	35.35	3.149	0.000000
CD	0.0113	2	0.0057	27.35	3.149	0.000000
ABC	0.0001	1	0.0001	0.57	3.958	0.476704
ABD	0.0003	2	0.0001	0.61	3.149	0.567456
ACD	0.0014	2	0.0007	3.46	3.149	0.032740
BCD	0.0049	2	0.0024	11.82	3.149	0.000015
ABCD	-0.0001	2	0.0000	-0.17	3.149	0.816789
ERROR	0.0099	48				
TOTAL	0.2110	71				

CO 8.2898 MSE 0.0002

ANOVA TABLE WITH 4 FACTORS:
A = 2 METHODS (PDM VS DDRR)
B = 2 DATA SETS
C = 2 MODEL VERSIONS
D = 3 OBSERVER TYPES

TABLE XX

ANOVA TABLE FOR TS1 - PDM VS DDRR

SOURCE	SUMSQ	DF	MS	FOBS	FCRIT	PVALUE
A	0.0004	1	0.0004	0.90	3.958	0.370369
B	0.2717	1	0.2717	615.67	3.958	0.000000
C	0.1454	1	0.1454	329.46	3.958	0.000000
D	0.4034	2	0.2017	457.12	3.149	0.000000
AB	0.0000	1	0.0000	-0.03	3.958	0.769925
AC	0.0000	1	0.0000	0.03	3.958	0.769763
AD	-0.0002	2	-0.0001	-0.24	3.149	0.775460
BC	0.0002	1	0.0002	0.45	3.958	0.522512
BD	0.0313	2	0.0157	35.48	3.149	0.000000
CD	0.0307	2	0.0154	34.81	3.149	0.000000
ABC	-0.0001	1	-0.0001	-0.14	3.958	0.683467
ABD	0.0001	2	0.0000	0.10	3.149	0.862142
ACD	0.0009	2	0.0004	1.00	3.149	0.399339
BCD	0.0005	2	0.0003	0.57	3.149	0.589589
ABCD	0.0001	2	0.0001	0.14	3.149	0.838830
ERROR	0.0212	48				
TOTAL	0.9055	71				

CO 20.4018 MSE 0.0004

ANOVA TABLE WITH 4 FACTORS:
A = 2 METHODS (PDM VS DDRR)
B = 2 DATA SETS
C = 2 MODEL VERSIONS
D = 3 OBSERVER TYPES

TABLE XXI

ANOVA TABLE FOR A0 - DISC VS DDRR

SOURCE	SUMSQ	DF	MS	FOBS	FCRIT	PVALUE
A	0.0003	1	0.0003	1.39	3.958	0.258223
B	0.2279	1	0.2279	1037.38	3.958	0.000000
C	0.0999	1	0.0999	454.85	3.958	0.000000
D	0.2543	2	0.1272	578.92	3.149	0.000000
AB	0.0002	1	0.0002	0.69	3.958	0.432009
AC	0.0001	1	0.0001	0.28	3.958	0.601672
AD	-0.0001	2	0.0000	-0.21	3.149	0.796003
BC	0.0000	1	0.0000	-0.07	3.958	0.735525
BD	0.0154	2	0.0077	34.98	3.149	0.000000
CD	0.0150	2	0.0075	34.14	3.149	0.000000
ABC	-0.0002	1	-0.0002	-0.97	3.958	0.347797
ABD	0.0000	2	0.0000	0.03	3.149	0.914440
ACD	0.0002	2	0.0001	0.38	3.149	0.690864
BCD	0.0006	2	0.0003	1.46	3.149	0.257405
ABCD	0.0003	2	0.0002	0.76	3.149	0.497225
ERROR	0.0105	48				
TOTAL	0.6244	71				

CO 25.9328 MSE 0.0002

ANOVA TABLE WITH 4 FACTORS:
A = 2 METHODS (DISC VS DDRR)
B = 2 DATA SETS
C = 2 MODEL VERSIONS
D = 3 OBSERVER TYPES

TABLE XXII

ANOVA TABLE FOR A1 - DISC VS DDRR

SOURCE	SUMSQ	DF	MS	FOBS	FCRIT	PVALUE
A	0.0001	1	0.0001	0.55	3.958	0.483529
B	0.0848	1	0.0848	390.08	3.958	0.000000
C	0.0448	1	0.0448	205.99	3.958	0.000000
D	0.0301	2	0.0151	69.29	3.149	0.000000
AB	-0.0001	1	-0.0001	-0.44	3.958	0.527387
AC	0.0003	1	0.0003	1.43	3.958	0.249962
AD	0.0004	2	0.0002	0.88	3.149	0.446950
BC	0.0044	1	0.0044	20.15	3.958	0.000004
BD	0.0131	2	0.0065	30.07	3.149	0.000000
CD	0.0087	2	0.0044	20.06	3.149	0.000000
ABC	0.0003	1	0.0003	1.30	3.958	0.275119
ABD	0.0001	2	0.0001	0.33	3.149	0.721399
ACD	0.0006	2	0.0003	1.49	3.149	0.248804
BCD	0.0028	2	0.0014	6.37	3.149	0.001721
ABCD	0.0002	2	0.0001	0.42	3.149	0.669063
ERROR	0.0104	48				
TOTAL	0.2011	71				

CO 8.1365 MSE 0.0002

ANOVA TABLE WITH 4 FACTORS:
A = 2 METHODS (DISC VS DDRR)
B = 2 DATA SETS
C = 2 MODEL VERSIONS
D = 3 OBSERVER TYPES

TABLE XXIII

ANOVA TABLE FOR TS1 - DISC VS DDRR

SOURCE	SUMSQ	DF	MS	FOBS	FCRIT	PVALUE
A	0.0002	1	0.0002	0.46	3.958	0.518262
B	0.2773	1	0.2773	642.89	3.958	0.000000
C	0.1484	1	0.1484	344.06	3.958	0.000000
D	0.3934	2	0.1967	455.98	3.149	0.000000
AB	0.0001	1	0.0001	0.18	3.958	0.658211
AC	0.0000	1	0.0000	-0.07	3.958	0.734412
AD	-0.0002	2	-0.0001	-0.18	3.149	0.815617
BC	0.0000	1	0.0000	0.07	3.958	0.734111
BD	0.0284	2	0.0142	32.86	3.149	0.000000
CD	0.0259	2	0.0130	30.05	3.149	0.000000
ABC	0.0000	1	0.0000	-0.11	3.958	0.706183
ABD	0.0001	2	0.0001	0.14	3.149	0.836747
ACD	0.0003	2	0.0001	0.32	3.149	0.727554
BCD	0.0002	2	0.0001	0.23	3.149	0.780631
ABCD	0.0002	2	0.0001	0.18	3.149	0.813851
ERROR	0.0207	48				
TOTAL	0.8950	71				

CO 20.5512 MSE 0.0004

ANOVA TABLE WITH 4 FACTORS:
A = 2 METHODS (DISC VS DDRR)
B = 2 DATA SETS
C = 2 MODEL VERSIONS
D = 3 OBSERVER TYPES

LIST OF REFERENCES

Box, George E.P., William G. Hunter and J. Stuart Hunter. Statistics for Experimenters. New York: John Wiley & Sons, Inc. 1978.

Box, George E.P., and M.E. Muller, 1958: A Note on the Generation of Random Normal Deviates. Annals of Mathematical Statistics 29, pp. 610-611.

Brier, G.W., 1950: Verification of Forecasts Expressed in Terms of Probability. Monthly Weather Review, 78, pp. 1-3.

Diunizio, Mark, 1984: An Evaluation of Discretized Conditional Probability and Linear Regression Threshold Techniques in Model Output Statistics Forecasting of Visibility Over the North Atlantic Ocean. Masters Thesis (R.J. Renard, advisor), Department of Meteorology, Naval Postgraduate School, Monterey, CA 200 pp.

Elias, Kristine C., 1985: Forecasting Atmospheric Visibility Over the Summer North Atlantic Using the Principal Discriminant Method. Masters Thesis (R.J. Renard, advisor), Department of Meteorology, Naval Postgraduate School, Monterey, CA 112 pp.

Glahn, Harry R. and Dale A. Lowry, 1972: The Use of Model Output Statistics (MOS) in Objective Weather Forecasting. J. Appl. Meteor., 11, pp. 1203-1211.

Karl, Michael L., 1984: Experiments in Forecasting Atmospheric Marine Horizontal Visibility Using Model Output Statistics with Conditional Probabilities of Discretized Parameters. Masters Thesis (R.J. Renard, advisor), Department of Meteorology, Naval Postgraduate School, Monterey, CA 165 pp.

Koziara, M.C., R.J. Renard and W.J. Thompson, 1983: Estimating Marine Fog Probability Using a Model Output Statistics Scheme. Monthly Weather Review, 111, pp. 2333-2340.

Lowe, P., 1984: The Use of Decision Theory for Determining Thresholds for Categorical Forecasts. Unpublished manuscript, Naval Environmental Prediction Research Facility, Monterey, CA 20 pp.

Preisendorfer, R.W., 1983a: Proposed Studies of Some Basic Marine Atmospheric Visibility Prediction Schemes Using Model Output Statistics. Unpublished manuscript, Department of Meteorology, Naval Postgraduate School, Monterey, CA 28 pp.

__________, 1983b: Maximum-Probability and Natural-Regression Prediction Strategies. Unpublished manuscript, Department of Meteorology, Naval Postgraduate School, Monterey, CA 10 pp.

__________, 1984: The Principal Discriminant Method of Prediction, Unpublished manuscript, Pacific Marine Environmental Laboratory, NOAA, Seattle, WA 25 pp.

__________, 1985: Notes on the Design of Simulation Data Sets for MOS (Model Output Statistics) Prediction Methods. Unpublished manuscript, Naval Postgraduate School, Monterey, CA 45 pp.

Renard, R.J., and W.J. Thompson, 1984: Estimating visibility over the North Pacific Ocean Using Model Output Statistics. National Weather Digest, 9, No. 2, pp. 18-25.

University of California, 1983: BMDP Statistical Software, 1983 Edition, Department of Biomathematics , University of California at Los Angeles, University of California Press, 726 pp.

Wooster, Michael H., 1984: An Evaluation of Discretized Conditional Probability and Linear Regression Threshold Techniques in Model Output Statistics Forecasting of Cloud Amount and Ceiling Over the North Atlantic Ocean. Masters Thesis (R.J. Renard, advisor), Department of Meteorology, Naval Postgraduate School, Monterey, CA, 187 pp.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Technical Information Center Cameron Station Alexandria, VA 22304-6145	2
2.	Library, Code 0142 Naval Postgraduate School Monterey, CA 93943	2
3.	Meteorology Reference Center, Code 63 Department of Meteorology Naval Postgraduate School Monterey, CA 93943	1
4.	Professor Robert J. Renard, Code 63Rd Chairman, Department of Meteorology Naval Postgraduate School Monterey, CA 93943	6
5.	Chairman (Code 68Mr) Department of Oceanography Naval Postgraduate School Monterey, CA 93943	1
6.	Dr. Rudolph W. Preisendorfer NOAA/PMEL/R/E/PM BIN C15700 Bldg. 3 7600 Sand Point Way, N.E. Seattle, WA 98115-0070	2
7.	Mr. Paul Lowe Naval Environmental Prediction Research Facility Monterey, CA 93943	1
8.	Mr. Robert Godfrey Naval Environmental Prediction Research Facility Monterey, CA 93943	1
9.	Director Naval Oceanography Division Naval Observatory 34th and Massachusetts Avenue NW Washington, DC 20390	1
10.	Commander Naval Oceanography Command NSTL Station Bay St. Louis, MS 39522	1
11.	Commanding Officer Naval Oceanographic Office NSTL Station Bay St. Louis, MS 39522	1
12.	Commanding Officer Fleet Numerical Oceanography Center Monterey, CA 93943	1
13.	Commanding Officer Naval Ocean Research and Development Activity	1

NSTL Station
Bay St. Louis, MS 39522

14.	Commanding Officer Naval Environmental Prediction Research Facility Monterey, CA 93943	1
15.	Chairman, Oceanography Department U.S. Naval Academy Annapolis, MD 21402	1
16.	Chief of Naval Research 800 N. Quincy Street Arlington, VA 22217	1
17.	Office of Naval Research (Code 480) Naval Ocean Research and Development Activity NSTL Station Bay St. Louis, MS 39522	1
18.	Commander (Air-330G) Naval Air Systems Command Washington, D.C. 20361	1
19.	Chief, Ocean Services Division National Oceanic And Atmospheric Administration 8060 Thirteenth Street Silver Spring, MD 20910	1
20.	Dr. Alan Weinstein, Code 422 Leader, Ocean Services Division Office of Naval Research Arlington, VA 22217	1
21.	Mr. Gil Ross, Met 09 Meteorological Office Bracknell, Berkshire England	1
22.	Chief, Technical Procedures Branch Meteorological Services Division National Oceanic And Atmospheric Administration 8060 Thirteenth Street Silver Spring, MD 20910	1
23.	Chief, Technical Services Division United States Air Force Air Weather Service (MAC) AirForce Global Weather Central Offutt AFB, NB 68113	1
24.	Steve J. Fatjo, LT/USN Naval Oceanography Command Center/ Joint Typhoon Warning Center, Guam Box 12, COMNAVMARIANAS FPO San Francisco, CA 96630-2926	2

END

DTIC

6-86